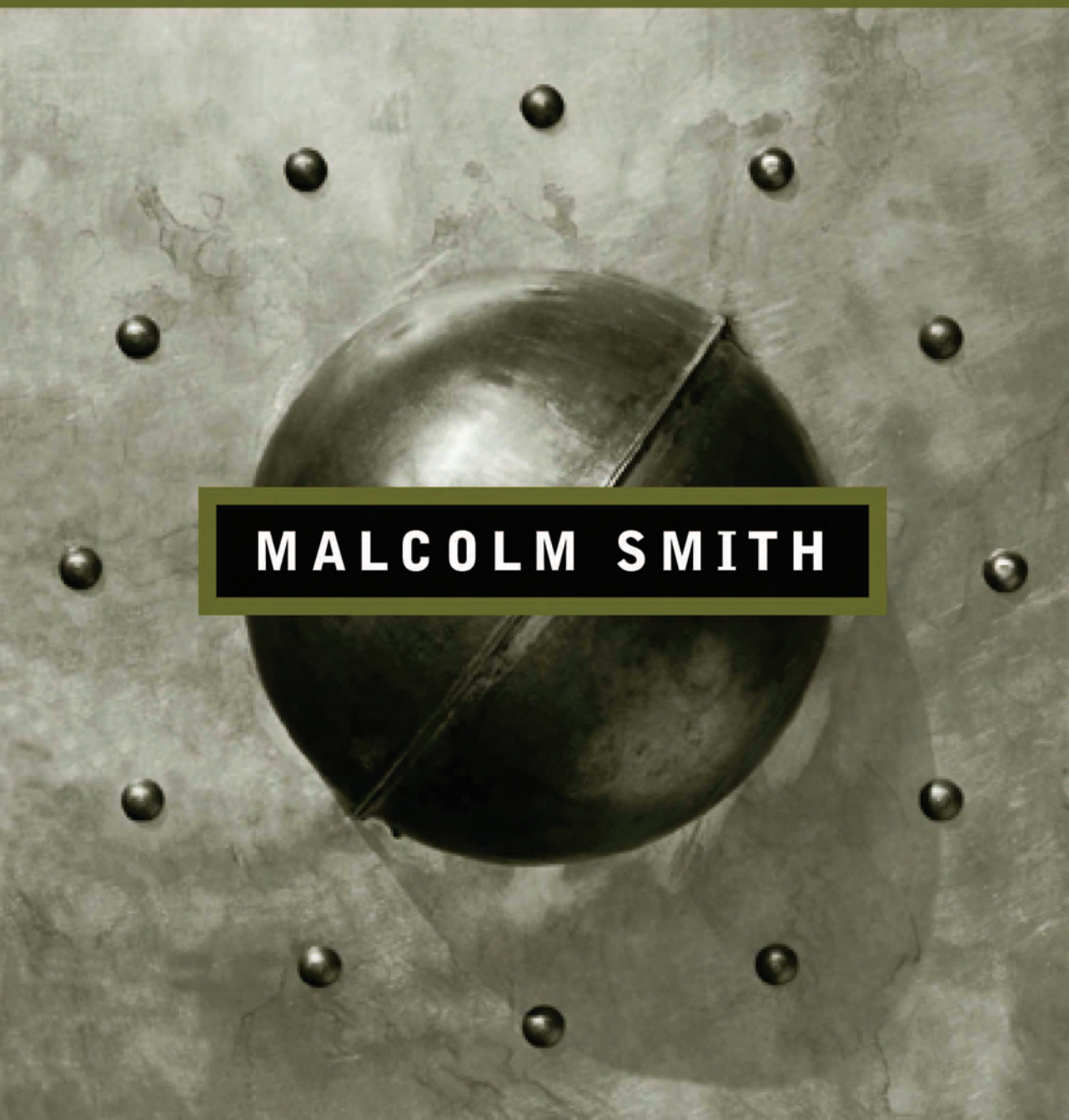
RESEARCH METHODS IN ACCOUNTING



MALCOLM SMITH

RESEARCH METHODS IN ACCOUNTING

MALCOLM SMITH

SAGE Publications
London • Thousand Oaks • New Delhi

First published 2003

SAGE Publications Ltd
6 Bonhill Street
London EC2A 4PU

SAGE Publications Inc
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B-42, Panchsheel Enclave
Post Box 4109
New Delhi 110 017

British Library Cataloguing in Publication data

A catalogue record for this book is available
from the British Library

ISBN 0 7619 7146 7
ISBN 0 7619 7147 5 (pbk)

Library of Congress Control Number available

Typeset by C&M Digitals (P) Ltd., Chennai, India
Printed in Great Britain by The Cromwell Press Ltd, Trowbridge, Wiltshire

This book is dedicated to Beth, Cedric and Alice

Contents

List of Figures

List of Tables

Acknowledgements

This book could not have been written without the help of numerous colleagues, most notably Bev Schutt and Paul Martin from the University of South Australia. Thanks are also owed to Glen Lehman, Linley Hartmann, Bruce Brown and Bruce Gurd at UniSA, and to David Russell, Ashok Patel and Elaine Harris at the Leicester Business School. However, as is normal, all errors and omissions remain the responsibility of the author.

Preface

This book aims to provide an insider's view of the research process, by focusing on actual choices made in the conduct of accounting research projects, together with a realistic perception of what might go wrong, even with careful planning. We must, however, acknowledge that no single author can be an expert in all research methods; this author is no exception. My own publications will readily reveal a preponderance of studies concerning experimental methods and the use of archival data; there are fewer instances of studies using survey and field study methods. It would be unwise of me to claim expertise in the implementation of all such methods, so this book must necessarily lean heavily on the work of others. For the same reasons, and because of pressure of space, this book does not address issues of finance, capital markets research, or stock-price-related accounting research on the fringes of finance.

Most other texts in this area are long, over-theoretical and not particularly user-friendly. This book aims to address these issues by adopting a practical approach which takes the reader from the initiation of the research idea right through to the publication of the research findings. The intended readership is wide, embracing instructors, doctoral candidates, and academics starting, or re-starting, their research careers. Although the focus, and examples, are mainly accounting based, much of the material will also be relevant to more general business applications, of particular interest to those pursuing a Doctorate of Business Administration (DBA) qualification. The practical examples employed are usually UK or Australia-based, these being the two countries in which I have extensive, current experience of teaching, supervision and examining, but the principles should normally adapt easily to alternative environments.

An early distinction between 'methods' and 'methodologies' in research is essential because the two are so often confused, or else used interchangeably. Research methods are concerned with the technical issues associated with the conduct of research; research methodology is concerned with the philosophies associated with the choice of research method. This book is almost exclusively concerned with the former and, after Chapter 1, deliberately neglects the philosophical foundations of research except where reference thereto is unavoidable.

Chapter 2 examines the research idea, and the documentary sources which might aid their development. A number of examples, many from non-accounting environments, are used to illustrate the research sequence, and to examine research that is seeking either to improve outcomes, to explain improved outcomes through new theory, or to examine the improvement

process itself. Theoretical frameworks and research models are used extensively here to help the reader to picture the key variables and relationships underlying their research.

Theory is the focus of Chapter 3, on the basis that 'good research is founded on good theory'. The chapter addresses the sources of the theory widely applied in accounting research, but drawn from other disciplines. In the space available it can only hope to give a flavour of the diversity which is available; indeed, it prompts us to suggest that 'theory in accounting research' might provide a suitable follow-up text in its own right! Recognition of the importance of theory, reliability and validity as desirable characteristics of accounting research lead, in Chapter 4, to the issues of data collection, management and analysis necessary to conduct hypothesis testing. This chapter is unashamedly quantitative in nature, but the relative strengths of qualitative analysis are addressed in subsequent chapters.

Chapter 5 addresses the increasingly important ethical considerations which underpin the conduct of accounting research, and the subsequent publication of research findings. It highlights the confusion which is still apparent among many academics as to what constitutes unethical conduct, and specifies the necessary guidelines for good practice.

Chapters 6 to 9 are devoted, respectively, to the core forms of accounting research: experimental, survey-based, fieldwork and archival. Numerous examples are used to demonstrate the relative advantages of alternative methods so that researchers can both make an informed choice and justify their preferred approach. Research can be based on quantitative or qualitative methods, and both should be equally acceptable as long as the most appropriate method has been chosen. Richardson (1996, p. 4) notes that 'work on how science really gets done' (such as that described in Chapter 2 based on Watson, 1968) shows that even though we are in an extreme 'positivist' domain, interpretive knowledge is still important in the development of new theory.

The majority of the readers of this text will likely be doctoral candidates so Chapter 10 is devoted to supervisor–candidate relationships, highlighting the mutual responsibilities of both parties to the supervision process, from the outset right up to the examination process.

Publication is the natural target output of the research process, and Chapter 11 addresses the complexity of the publication process. In doing so it recognises that we are working in a dynamic process; what was once acceptable in accounting research is no longer so because of a more appropriate emphasis on research ethics; what is publishable, at all or in specific journals, changes too, both with the passage of time and the passing of particular journal editors. Many journals remain very conservative in the type of research they will publish, often on the grounds that it is difficult to demonstrate that 'new' methods constitute 'good' research in the same way as the traditional methods. But this situation is changing gradually – the wider opportunities for publishing case-based research in recent years provides evidence of this. However, the renewed emphasis on journal and university rankings, and associated funding systems based on the quality of publications, provides fresh difficulties.

The provision of 'acceptable journals' listings by many universities, and the prohibition of publication elsewhere, perpetuates the position of the well-established journals, while making it extremely difficult for the editors of other journals to attract quality submissions. The opportunities for innovative new journals are also severely diminished in such circumstances.

Contributions to the profession by academic accountants are generally not well regarded, either by one's colleagues or by government bodies providing funding based on publications performance, even though, arguably, the education of the potential employers of our students might be seen as an important part of our jobs. So journalistic pieces in practitioner magazines and workshops to professional audiences count for close to nought – even though the individuals concerned would never read a refereed journal or attend an academic conference. We need to exploit the available media to get our message, and the power of research findings, over to those implementing change in an unbiased way, before the consultants get in on the act! This process must be of mutual benefit to all parties, but if the practitioners feel they are being short-changed, or even used, then future collaborative efforts will be threatened. It is just such attitudes which generate the 'them and us' cultures leading to accusations of academics being out of touch with reality. In this context it is interesting to note the changes taking place within professional journals: there were once two such journals called *Management Accounting*, but now there are none, the US version becoming *Strategic Finance* and the UK one *Financial Management*. With moves to term 'accounting' as 'assurance services' it will surely not be long before some of the professional bodies themselves follow their journals with the removal of the word 'accounting' from their titles.

Communication problems also remain. The timeliness and relevance of much of the content of the refereed literature does little more than suggest that it is written by academics, only for the consumption of other academics! Most practitioners do not have an appreciation of research methods, nor do they read the refereed literature, so important findings and recommendations often do not reach the individuals who can make sure it has the greatest impact. A number of journals have emerged with the express intention of providing readable research for practitioner audiences, but even these have tended to become more academic and less readable over relatively short time periods. This book aims to provide a treatment of research methods that will be of use to both accounting practitioners and those contemplating the conduct of research projects.

Space restrictions mean that this slim volume cannot hope to tackle all of the detail of the application of different research methods, or the associated intricacies of complex quantitative methods. But if its use causes one paper to be published that would otherwise have gone unpublished, then all will have been worthwhile.

Malcolm Smith

1

Introduction and Overview

CHAPTER CONTENTS

A number of authors (e.g., Brownell, 1995, p. 2) describe accounting researchers as 'parasites' who prey on the work of others to generate their findings. The term may be an overstatement, but as with most rash generalisations it contains more than a germ of truth: accounting researchers have little theory of their own (they rely on economics, finance, psychology, sociology and organisational behaviour as their major sources); they have no methods of their own (they are all adapted from the natural and social sciences); and they have few instruments of their own (with many of these originating in or adapted from the organisational behaviour literature). Merchant (quoted in Brownell, 1995, p. 140) even suggests that organisational behaviourists are much better at developing survey instruments than their accounting counterparts.

The overall aim of this book is to facilitate the conduct of applied research studies in accounting, and to do this we must recognise our reliance on work in other disciplines. To accomplish this aim, a number of subordinate objectives may be identified, all of which will contribute to the overall goal:

- an understanding of contemporary research ideas in accounting, so that readers can identify and define research problems and prepare strategies for their solution;
- an awareness of alternative research methods, to facilitate the selection of the most appropriate method for addressing particular research questions;
- an ability to review existing research and to offer critiques of articles published in refereed journals; and
- an appreciation of the ethical constraints on the conduct of accounting research.

Research in accounting is concerned with solving problems, investigating relationships and building a body of knowledge. Because we rely to such a great extent on prior research in the natural and social sciences to do so, this volume will take a similar approach in leaning on work in other disciplines where it helps to inform accounting research.

Bennett (1991) identifies four basic levels of research:

- Description – concerned with the collection and reporting of data related to what is, or was, the case. This would include means and standard deviations of individual variables, and correlations between pairs of variables.
- Classification – still descriptive, but easing the reporting process, and highlighting similarities and clustering through grouping and classifying (e.g., through the familiar cross-tabulation facility in most basic statistical packages).
- Explanation – an attempt to make sense of observations by explaining the relationships observed and attributing causality based on some appropriate theory.
- Prediction – going beyond the understanding and explaining of the prior stage, to model observations in a way that allows testable predictions to be made of unknown events.

We return to this structure in Chapter 4 when discussing alternative quantitative methods, but an early distinction between 'explanation' and 'prediction' is appropriate here, because, as in the natural sciences, we are able to make excellent predictions of accounting behaviour without the backing of a sound underpinning theory. Bankruptcy prediction modelling provides an excellent example. A number of researchers (e.g., Altman, 1968; Taffler, 1983) have developed models that have proved very successful in identifying 'distressed' companies – those companies that will fail in the short term. These models are statistically excellent but the theory underpinning their content, in terms of the ratios to be used and the variables they represent, is extremely weak; the essential problem is that such theories as we have (e.g., Blum, 1974; Myers, 1977; Scott, 1981) do not generate very good predictive models!

Good research generates the sound evidence needed to overturn or revise existing theories. These assertions will, in turn, yield to revised theories based on better evidence, so that healthy competition between rival ideas will lead to better explanations and more reliable predictions. Two major processes of reasoning, 'deductive' (theory to observation) and 'inductive' (observation to theory), are important for theory construction and observation testing. Inductive reasoning starts with specific observations (data) from which theories can be generated; a generalisable pattern may emerge from further observations and repeated testing for compliance. The natural sciences, for example astronomy, provide numerous examples of inductive reasoning, thus Hawking (1998) provides a number of fascinating examples of theories revised, or still in question, with implications for the progress of accounting research. However, he notes that generalisations made on the basis of induction

can never be regarded as 'certain', since just one contrary instance can cause them to be overturned:

- **Big Bang versus Steady State.** From the late 1940s to the mid-1960s two competing theories were prominent in offering alternative explanations of the origins of the universe. The 'Big Bang' theory recognised a singular event as causing an ever-expanding universe in which matter (notably galaxies) becomes continuously more widely dispersed. The 'Steady State' theory, attributed to Bondi, Gold and Hoyle, on the other hand, suggested that matter was continuously being created to fill the gaps between existing galaxies. They argued that the universe had no beginning, and had been forever expanding, with new matter being created out of apparently empty space. The Steady State theory importantly provided testable hypotheses in suggesting that the universe should look the same at all times and from wherever it was viewed. But surveys of radio waves in the early 1960s showed that sources were more numerous in the past, and that there were many more weak (distant) sources than strong (close) ones. Further, microwave radiation studies in 1965 demonstrated that the universe did not have a common density – it had been much denser in the past. These observations provided disconfirmations of the Steady State theory, causing its abandonment.

- **Newton's Laws of Physics.** New theory emerges when a new observation arises which does not correspond with existing theory. Once the technology permitted accurate observations of the planet Mercury to be made, it was clear that there were small differences between its observed motion and that expected under Newton's Theory of Gravity. Einstein's general theory of relativity matched the observed motions of the planet in a manner that Newton's theory did not, providing confirmation for the new theory.

- **The Wave Theory of Light.** We can attempt to explain the behaviour of light in terms of its being composed of either 'waves' or 'particles'. Each view produces a plausible explanation of behaviour – both of which are needed to affirm existing properties – but they are incompatible explanations which cannot exist simultaneously. New theories are required (possibly those associated with parallel universes) for a complete understanding of the incompatibility.

Deductive reasoning, on the other hand, starts with the theory and proceeds to generate specific predictions which follow from its application. The predictions can be verified, or otherwise, from subsequent observation. For example, in his seminal paper, Healy (1985) used agency theory to develop a bonus hypothesis which could be substantially verified through observations of how managers manipulated their accounting earnings to optimise their short-term bonus performance.

However, such a strict division of reasoning processes is not always helpful because interdependencies almost always exist: induction will usually imply some knowledge of theory in order to select the data to be observed (a common criticism of grounded theory advanced in Chapter 8); deduction will be dependent on the selection of the initial hypotheses for testing.

Even without such problems, the scientific position of 'objective measurement' has come under repeated attack, in both natural and social sciences, because the act of observation is itself 'theory-laden' and influenced by the motives and preferences of the observer. For example, Hopwood (1987), in management accounting, and Hines (1988), in financial accounting, argue that accounting helps to create the 'facts' that it is supposedly reporting. More radical approaches (e.g., Tinker and Niemark, 1987) suggest that accounting distorts practice in a systematic manner. Such concerns have aided the development of new approaches: an interpretive perspective and a critical perspective.

- **An interpretive perspective** – From an interpretive perspective, human actions are the result of external influences. These actions have both intentions and reflections, and take place within a structure of rules which binds the participants. The task of the researcher goes beyond measurement to developing an understanding of the situation. To do this effectively, active participation, rather than detached observation, may be required. Since the 'action' may be interpreted ambiguously when taken out of context, this perspective places the fundamental emphasis on the understanding of the process. In an accounting context, Arrington and Francis (1989) provide an example of this approach.

- **A critical perspective** – The critical approach expands on the scope of the interpretive approach by focusing on the ownership of knowledge and the associated social, economic and political implications. An empirical approach is criticised on the grounds that the research process is value-laden, and that the acquisition of knowledge provides the opportunity to oppress those being researched. In an accounting context, Tinker (1980) provides an example of this approach.

Table 1.1 summarises the differences in research assumptions, process and outcomes associated with each of these three major approaches.

Kuhn (1970) suggests that researchers are concerned with problem-solving within a single framework of widely accepted beliefs, values, assumptions and techniques. This shared framework, or view of the world, he termed a paradigm, so that a 'paradigm shift' corresponds with some revolution where the existing framework and theories can no longer cope with the volume of disconfirming evidence. Kuhn neatly illustrates such a shift by reference to a simple psychology experiment:

> Subjects viewed cards from a deck. The deck included some unusual cards, including black hearts and red spades, but the subjects were not informed in

TABLE 1.1

Three alternative approaches (adapted from Connole, 1993, p. 37)

Positivist	Interpretive	Critical
1. *What is the approach modelled on?* Classical investigation founded in the physical sciences.	Historical, literary and existential studies in which the subjective understandings of subjects are significant.	Marxist and interpretive studies which focus on the insights and judgements of the subjects.
2. *What does it assume about reality?* Reality is unitary and it can only be understood by empirical and analytic methods, i.e., the scientific approach.	There are multiple realities which require multiple methods for understanding them.	There are multiple realities which are made problematic through distorted communication.
3. *What is the foundation of data?* Disciplined rules for observation.	Meanings are the basis of data: meaning precedes logic and fact.	Meanings are found in language and social behaviour and they precede logic and fact.
4. *How is observation done?* Through clear and unambiguous rules which are not modified by the setting and are totally independent of it.	Through the social, linguistic and cognitive skills of the researcher.	Interpretive methods, plus critical self-reflection concerning the grounds of observation.
5. *What is generated?* Evidence and generalisable laws which are not affected by contexts and have nothing to do with the way in which they were discovered in the first place. Objectivity depends upon the removal of error and bias which is related specifically to the logic of observation and measurement.	Knowledge which is dependent on the process of discovery. The integrity of the findings depends upon the quality of the social, linguistic and cognitive skills of the researcher in the production of data analyses and conclusions.	Knowledge which falls within the interpretive framework, but which also serves the purposes of assisting personal liberation and understanding, and emancipation from forces constraining the rational independence of individuals.
6. *What interests are inherent?* Prediction and control, technically exploitable knowledge, and explanation.	Understanding at the level of ordinary language and action. Discovering the meanings and beliefs underlying the actions of others.	Interpretive interests and those which underliey other forms of inquiry. Radically improving human existence. Practical and public involvement in knowledge formation and use.
7. *What values are inherent?* Science and scientific knowledge are inherently value-neutral.	Science and scientific knowledge have both to be interpreted in terms of values they represent.	Science and knowledge are never value-neutral: they always represent certain interests.

advance about their presence. Initially the subjects saw only 'hearts' and 'spades', because they believed that only 'red hearts' and 'black spades' existed; only with repeated viewing did they grasp that these cards were not typical of a normal deck. Then they could recognise the cards that existed rather than the ones they were expecting.

In accounting research the parallels might be the paradigm shifts associated with the ideas introduced by Ball and Brown (1968) and the difficulty they had in getting a paper published which questioned the existing paradigm by showing a link between stock prices and accounting earnings, through the abnormal performance index. A similar, though perhaps less radical, movement is associated with Watts and Zimmerman (1978) and their popularisation of agency theory in an accounting environment.

What is inescapable is that we are dealing with people, and in the research community that means individuals with their own agenda and with reputations to build and protect. The natural sciences are littered with character assassinations of individuals and their work, by others who have been less than willing to accept the impact of new findings on their own fiefdoms.

Sir Humphrey Appleby, in Lynn and Jay (1987), outlines the four stages of the process necessary to discredit an unwelcome report. The parallels between the fictitious Department of Public Administration and academia are uncomfortable, where unwelcome findings might arise from academic competitors:

1. **Refuse to accept the findings** on the basis that they could be misinterpreted, and that a wider and more detailed study is required.
2. **Discredit the evidence** on the basis that it is inconclusive and the figures are open to other interpretations, or that the findings are contradictory and leave important questions unanswered.
3. **Undermine the recommendations** because they say nothing new, and provide insufficient information on which to draw valid conclusions.
4. **Discredit the researcher** by questioning his or her integrity, competence and methods employed.

We thus have doubts about the researchers, their research questions, their research methods, the means of data collection and analysis, and the validity of the interpretation and recommendations – all issues to which we will return.

Theory as testable explanation

Faced with a set of diverse observations, we can establish a set of tentative explanations which help to make sense of the diversity. Such explanations

constitute theory. In any set of circumstances there will usually be multiple theories available to explain the observations. The systematic collection of further data allows for the testing of the alternative theories so that we can establish which of the exiting theories best explains the facts. A layman's perspective of 'theory' is cynically expressed in Michael Crichton's *The Lost World* as: 'A theory is nothing more than a substitute for experience put forward by someone who does not know what they are talking about' (1995, p. 67).

The data collection itself allows only a descriptive approach (e.g., means, standard deviations, ranges, correlations); we cannot attempt to attribute causation in any meaningful way without recourse to an explanatory theory. We are always looking for another theory which may fit better, so that, as Popper (1959, p. 104) suggests, a 'genuine test of a theory is an attempt to falsify it or refute it'. We look for disconfirmations rather than confirmations.

In the short term this may not be successful. In accounting, we witness the frequent and numerous 'anomalies' to which the Efficient Markets Hypothesis (EMH) is subject, but we have no other widely accepted theory of the manner in which stock prices react to the availability of relevant information.

Popper's suggestions are very attractive in providing a powerful empirical methodology for subjecting theories to attempts to refute them. However, this position is not always ideal because the process of 'observation' in itself may be fallible. Thus Hawking (1998) reports *Heisenberg's Uncertainty Principle*:

> If we are to predict the future position and speed of any particle, then we require accurate measurement of both its present position and current speed. Heisenberg did this in 1926 by shining light on a particle, and observing the resultant scattering of light in order to reveal its position. However, to determine the position of the particle accurately an amount of light needed to be used which changed the speed of the particle in an unpredictable way: the more accurately he tried to measure position, the less accurate was the measurement of speed!

The Uncertainty Principle has wide implications for research conducted in any environment, where it is impossible to measure the size and speed of a particle without altering all other characteristics in the process of measurement. We have a parallel situation in accounting research where the actions of the participants in ethnographic, experimental, survey or fieldwork impacts on the outcomes of the measurement process.

Three fundamental criteria exist to judge whether theory fits observation:

1. **Co-variation** – even where no causality exists we would expect the two variables to move together so that a high degree of correlation exists between the two variables. Where there is no co-variation it will be difficult to establish a causal link.

2. **Cause prior to effect** – if a causal link is to be established then the 'causal event' should occur before the 'effect event'. The sequence of events can therefore help to establish an explanatory direction.
3. **Absence of plausible rival hypotheses** – the third rule seeks to eliminate alternative explanations of the events as being implausible. This may only be possible in the present, because future researchers may develop competing explanations of the events from a re-analysis of the data.

Consider, for example, the voluntary disclosure of information in corporate reports and analyst following (i.e., the number of analysts examining the performance and reporting on the disclosures of large companies). There is a relationship between these two variables – they co-vary: the volume of voluntary disclosures and the number of analysts reporting move together. But which is causing which? Rival hypotheses suggest:

a. companies are supplying more information voluntarily to the market to signal their intentions and reputation, attracting the attention of more investment analysis;
b. investment analysts are focusing their attention on particular companies and demanding more information and more detailed disclosures.

The existing empirical evidence is less than convincing: Lang and Lundholm (1996) find (a); but Walker and Tsalta (2001) provide only weak evidence for (a) but stronger evidence to support (b). Clearly more empirical work is required to clarify the direction of causation.

A critical approach to accounting research

Researchers must demonstrate a healthy scepticism towards both their own findings and those of other researchers. They must adopt a critical posture, questioning everything that they read until sufficient evidence has been provided for them to be satisfied with the quality of the outcomes. The development of critical appraisal skills is a fundamental requirement in researchers, so that they can distinguish between good and bad research, and clearly identify flaws of argument, methodology and analysis.

Honest and transparent reporting of research practice is an ethical duty of those participating. Researchers should report everything that they did, why they did it, and how they did it. If they have doubts about any stage of the procedure, then these should be stated, along with their likely implications and what, if anything, has been done to overcome these doubts. Where researchers have been 'economical with the truth', this is usually apparent in their papers and is often an indicator of bad research.

Students frequently struggle initially when they are asked to critique published articles. They are often in awe of the reputation of the authors, or doubt whether they are able to offer sensible criticism of papers which, after all, have already undergone editorial scrutiny and double-blind review. Despite the above, some flawed papers do get published, and these are not always in lower-tier journals (see Hartmann and Moers, 1999, for their critique of 28 papers on contingency analysis in three top accounting journals – *Accounting Organizations and Society* (AOS), *The Accounting Review* (AR) and *Journal of Accounting Research* (JAR) – in which they identify problems in the design and analysis of 27 of the studies!). With appropriate guidelines as to the appropriate questions to ask, students can quickly develop some confidence in their ability to spot flaws and omissions. For example, Abernethy, et al. (1999) provide a stimulating critique of the three subsequent papers in the same, outstanding, edition of the journal *Accounting & Finance*.

We would usually want to address the following:

1. **Why is this article interesting/important?** The paper must offer some new insights which constitute a contribution to knowledge. These insights should be non-trivial, so that they can be embraced in either further theory development or recommendations for improvement.

2. **Are the outcomes important?** Effectively, does the paper pass the 'so what' test? Will anyone be interested in the outcomes of this research, or will it have any implications for future practice? Would the scope of the research be well regarded by competitive grant authorities? This has important implications for those papers which produce 'negative' findings, that is, they test reasonable hypotheses based on the research literature, but their datasets fail to support any of the expectations. These findings still make a contribution in that they demonstrate that findings from elsewhere (often other disciplines) do not hold in accounting, but their negativity may restrict their publication opportunities.

3. **What motivates the authors to write this article now?** The paper may be clearly addressing issues of contemporary concern; on the other hand, it may be addressing more historical issues and/or be using 'old' data. If we have the latter, we may be dealing with an old paper recently recycled, or a paper which has been through many iterations at several different journals before being deemed 'publishable'.

4. **What is the research problem/question?** We are looking for a clear statement of the problem very early on in the paper, so that its objectives are readily apparent. If we reach page 11, say, of the paper without a clear idea of its direction, or any sort of research model, then perhaps the authors need to readdress the fundamental purpose of the research.

5 **What theory or theoretical framework underpins the research?** Without some theoretical foundations we have a problem-solving exercise or a consultancy project, neither of which should be gracing the pages of a refereed journal. There must be some theoretical justification for the question being addressed and the research approach adopted. Theory will often not come first in the research process – it will frequently be preceded by an interesting idea or a perplexing observation. But we require some theoretical explanation for the relationships under investigation before we have the basics of a refereed journal article. Observed deficiencies in this area usually fall into one of four categories:

- the underlying theory is either non-existent or extremely thin;
- the theoretical context is there but appears to have been tacked on as an afterthought – usually at the beginning of the paper and often written by a co-author. Examination of writing styles suggests that we frequently do not have a seamless divide between 'theory' and 'conduct of research';
- the theoretical arguments are unconvincing, so that there are competing theories that may reasonably have been adopted in the paper but have been overlooked;
- a sound theoretical framework but findings which are totally at odds with theory. Apparently, a competing theory may be more appropriate, although this is unknown to the authors at the time.

6 **What are the key motivating literatures on which the study depends?** There will normally be a small number of seminal pieces of literature which are driving the research. If any of these are themselves unreliable, it may cast doubt on the state of the foundations on which the paper is based. If one of the papers is an unpublished conference or working paper from several years before, then alarm bells ring to question why that piece has not itself been published in the refereed literature. If key seminal pieces of literature have been overlooked, then again the integrity of the findings is reduced.

7 **Which research method has been chosen?** There should be a justification for the chosen method, and a clear preference over alternatives. The method should be consistent with both theory and literature and, ideally, prior empirical studies in the field will have adopted similar methods. Most importantly, we want to see a research method that has evolved rather than one that has been selected first, even before the research question has been fully developed. The use of survey methods should always be questioned in this way since frequently they seem to have been selected without explanation of the elimination of alternatives. Ideally, we should be able to trace through the emergence of abstract concepts, from theory, through their operationalisation and measurement, so that any hypotheses are entirely consistent with both theory and literature.

8 **How has the sample been selected?** Details on sample selection are often sketchy in many articles, perhaps because the authors feel vulnerable about the procedures adopted. Sometimes (see, for example, Young, 1996) the actual sample size employed is omitted, as is the response rate. Both omissions should be regarded as bad news. It is usually clear that scientific methods have not been adopted (unfortunately far too commonly in accounting research) where there is an over-reliance on convenience samples. What may be apparent is an attempt by the authors to obfuscate in this regard, to overlook detail and try to create an impression that the sample selection is more systematic than it has actually been.

9 **How have questions of validity been addressed?** Choice of research method should address issues of validity. Where experimental methods have been employed we would anticipate questions of internal validity to be paramount; where field studies are involved we would expect issues of external validity to be addressed. For survey methods we would anticipate the focus to be more on the reliability of the test instrument and the rigour of the subsequent statistical analysis, rather than on validity issues.

10 **How have the results been analysed?** We want to see the simplest analysis of the results consistent with the relationships being explored. We do not wish to see unnecessary complexity; this will make the paper less readable and tend to mask the findings and their significance. On the other hand, most academic accountants are only 'amateur' statisticians; if the level of their analysis is inadequate, then they may need to bring in a statistician as co-author (evidenced by the number of 'quant jocks' appearing as third or fourth authors on accounting papers to satisfy the reviewers). Importantly, we do not wish to see the method of analysis driving the study. In just the same way as the research method should not precede the research question, then neither should the method of analysis. For example, I recall a paper of my own (M. Smith, 1992) presented at a conference but never published. It attempted to show the advantages of using multidimensional scaling (MDS, then a little-used technique in the accounting literature) for problem-solving, but the journal referees rightly observed that the method was inappropriately sophisticated for a relatively simple research question. MDS was abandoned, simpler methods instituted and the revised paper eventually published as Smith (1996).

11 **Are the conclusions and recommendations consistent with the findings?** Effectively, does the paper hold together? Is the title appropriate? Do the abstract and introduction lead us to expect what we find at the end of the paper? In many papers the final sections are the weakest and may not do justice to the breadth of the research conducted. We look for explanations, limitations and a future research agenda.

Let us now consider how this framework may be applied to a critique of a published piece. Naturally, I chose one of my own publications (Smith et al., 2001) for the treatment because a knowledge of the history of the development of the paper, from an insider's perspective, can be most instructive. Readers will be able to make the most of the subsequent discussion if they are first able to read a copy of the paper, and for this purpose the complete paper is reproduced as Appendix 2.

1. **Interesting new insights:** The paper posits an interesting connection between (1) audit firm; (2) manner of conduct of the audit; and (3) classification of audit firms based on their procedures and culture. The paper also attempts to impose a global perspective by employing findings from the USA, the UK and Australia. But neither the data nor the supporting literature is new, and it compromises the originality of the paper.

2. **Importance:** The paper is important if it makes a contribution to knowledge. This may be a contribution to theory development or implications for business practice. If the paper can demonstrate a relationship between 'auditor' and the manner in which the audit has been conducted, then this makes a contribution, even though it may only be of historical relevance. Such a relationship is shown for 1987/88 data, but evidence is also presented to suggest that this relationship no longer holds. The absence of a current relationship suggests that the paper has no implications for current auditing practice. The reasons why a relationship between the audit firm and its propensity towards tolerance of particular accounting policies among its clients is by no means clear.

3. **Motivation:** The timing of the paper is problematic. It is published in 2001 but uses data predominantly from 1987/88. There is a danger of its being regarded as a historical piece with little relevance to current practice. The authors justify the use of this dataset in that the Kinney classification, the target test of the paper, is based on data relating to the Big 8 group of accountants, with 1988 being the last year of existence of the Big 8 in Australia, prior to extensive merger activity in the sector. There is the suggestion, though, both from the paper itself and the references cited, that the data have been used primarily to generate failure prediction models for the Western Australian government (i.e., Houghton and Smith, 1991) and that the further use of this data in this paper may be incidental and opportunistic.

4. **Problem statement:** The problem statement is quite clearly stated as: Accounting Policy Changes = f {auditing firm}, where both sides of this equation are elaborated and measured for a large number of companies:

   - Accounting policy changes: discretionary/mandatory; income increasing, income reducing, neutral.

- Auditing firm: by individual name, and by grouping according to classifications developed by Kinney (1986) and Moizer (1998).
- A number of extraneous variables (notably firm size, financial performance and industry) are also examined to determine their impact.

5 **Theoretical framework:** This remains something of a problem with this paper, despite strenuous efforts to overcome omissions. The literature demonstrates that there are differences between auditors, and in the procedures that they adopted for audit in 1988 (i.e., Cushing and Loebekke, 1986; Sullivan, 1984). However, why these procedural differences between auditors translate into differing tolerances towards income-impacting accounting policy changes is unclear, and is largely attributable to unpublished anecdotes from practising auditors and the discussion arising in a single paper (Dirsmith and Haskins, 1991).

6 **Motivating literatures:** Relatively few articles, noted above (i.e., Sullivan, Kinney, Cushing and Loebekke, Dirsmith and Haskins) motivate this paper, while Terry Smith (1992) and Peter Moizer (1998) provide the opportunity for UK comparisons. The pivotal paper is Dirsmith and Haskins (1991), published after the conduct of the data collection; there is thus a strong suspicion that interesting findings have arisen from data mining operations in 1988, for which Kinney (1986) provides a conceptual framework, but that publication must wait for a suitable theory. There is very little other supporting literature, though self-citation by the authors is also revealing:

- Houghton and Smith (1991) relates to failure prediction models constructed with the same data and is employed here to provide a measure of overall financial performance;
- Smith (1998a) reports current UK findings linking auditor with attitude to accounting policy change;
- Smith and Kestel (1999) update the present study with a time series analysis, but the results are apparently insufficiently interesting to constitute publication in a refereed journal;
- Brown (1988) reports on the most appropriate means of conducting statistical tests with contingency tables.

7 **Research method:** Archival methods are employed, since they are the only realistic alternative given the nature of the data: namely, historical, documentary, and covering many companies that are no longer in existence. The authors' access to a dataset comprising the population of Western Australian public companies is a considerable strength of the paper. Data collection is meticulous and involves checks for consistency both between individual researchers and for temporal validity.

8 **Sample selection:** The paper accesses the annual reports of all 463 publicly quoted companies in Western Australian (WA), so does not encounter any sampling issues other than a restriction on the nature of statistical tests that may be employed because of using a population rather than a sample from a normal distribution.

9 **Validity issues:** There are potential internal validity threats consequent upon the failure to consider competing theoretical explanations for the observations. The incidence of accounting policy change is apparently associated with auditing firm, but both the direction of causation for the relationship and alternative auditor motivations might be considered. The authors acknowledge the lack of external validity in the study – the applicability of the findings to other time periods and other datasets – in that conditions have changed so substantially since the data collection period that the procedures adopted by all auditors are now very similar.

10 **Analysis:** The fundamental analysis is relatively unsophisticated, involving the comparison of 'observed' and 'expected' frequencies through a chi-squared test. A variation on the traditional approach is introduced to take account of an ordering effect in the contingency tables, the power of the tests being increased with the use of Kendall's-tau. (A co-author with specialist statistical publications has been included to address testing issues, potentially in response to reviewer concerns on previous versions). A comparative fundamental analysis for UK data (alluded to in Smith, 1998a) is apparently not possible, and further analysis is restricted to tertiary sources.

11 **Conclusions:** There are no formal conclusions or recommendations, rather a discussion of other interesting findings in related fields which may impact on the integrity of the outcomes. The findings of this study are linked to merger activity in the Big 8, showing a pattern with considerable similarities to past successes. The paper suggests that future merger activity in the sector may be influenced by the organisational culture aspects of the Kinney classification and the clustering of companies generated by Moizer (1998); thus if we were looking at potential suitors for Arthur Andersen, say, then the analysis suggests that Ernst & Young would provide potentially the most successful alternative.

Such a critique is revealing, giving glimpses of a less-than-optimum approach adopted in the development of this particular paper. Data were collected for the specific purpose of generating failure prediction models for the WA government, and corporate monitoring of distressed enterprises (i.e., Houghton and Smith, 1991). The interesting auditor findings were generated at the same time, but there was no substantive theory to justify the observed relationship – and consequently no research paper. Only with the emergence of new theories (e.g., Dirsmith and Haskins, 1991), which might motivate the study, could further development towards a publishable paper proceed.

Clearly research is not always simple, systematic and clean – despite the sanitised versions that we read in the published journals. The research process can be both chaotic and exciting, and very rarely proceeds exactly according to plan. Unfortunately, this impression is rarely created by what we read because published pieces usually have happy endings – positive findings and co-operative participants. For a more realistic version of events we must rely on books like this, conference presentations and research workshops!

Armed with a critical and sceptical approach to the research of others, we can now start to develop the skills required to conduct competent research of our own, and commence a sequence which will eventually result in the publication of our research findings.

2

Developing the Research Idea

CHAPTER CONTENTS

We recognise in Chapter 1 that research processes are usually neither simple, systematic nor clean because research rarely proceeds exactly to plan. However, this should not deter us from planning thoroughly in the first place to specify how, in an ideal world, we would like the research to be conducted.

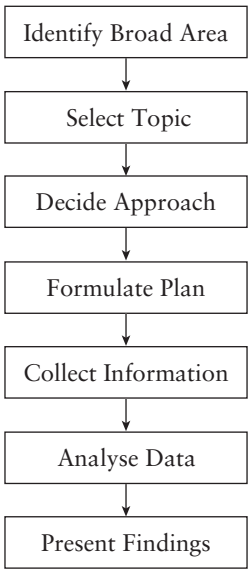
The research sequence

Figure 2.1 specifies the typical research sequence described by Howard and Sharp (1983) as a series of stages we would expect to progress through in most forms of accounting research, while moving from original idea to eventual publication.

The following chapters of this book address each of these stages, and detail the constraints we might anticipate.

1. **Identify broad area:** Narrow the focus from accounting in general to a stream associated with financial accounting, management accounting, auditing, accounting education or accounting information systems.

FIGURE 2.1



The research sequence

2. **Select topic:** Specification of a sub-area to provide a tighter focus, and one for which supervision capacity is available, but one which may be modified in the light of subsequent developments.

3. **Decide approach:** Early thoughts regarding the approach to be adopted will revolve around the resources available, and in particular access to the necessary data sources. A detailed specification of research methods to be adopted must wait until the literature review has been conducted and theoretical foundations and outline hypotheses have been established.

4. **Formulate plan:** Milestones and targets should be established at the outset so that it is clear how the research will progress over an extended period. This is particularly important for part-time researchers who may be contemplating study over six or seven years. A Gantt chart, or similar, is very helpful in clarifying the extent of the programme of work and the mutual expectations of those involved, especially if this concerns the relationship between candidate and supervisor in doctoral work. This plan should include target conferences where preliminary findings may be presented, especially where deadlines are important to the candidate. The commencement period can often cause anxiety because of the

perceived need to make swift progress. It cannot be emphasised enough how important an extended period of reading is to ensure that effort is not wasted performing experiments or surveys which subsequently emerge as being unnecessary or fatally flawed.

5 **Collect information:** Data collection can safely proceed only when we recognise exactly what we want to know, and for what purpose. The planning stage should highlight the period over which we want to collect data; this usually effectively precludes most longitudinal studies, partly because it takes too long to collect data and partly because of the increased vulnerability associated with extended site access. We may require access to commercial databases; if these are an essential requirement then permissions should be sought immediately.

6 **Analyse data:** Methods of data analysis and software requirements should be apparent early in the research process.

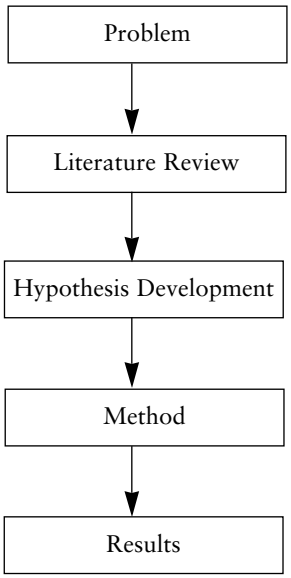
7 **Present findings:** Preliminary findings will normally be presented at university workshops and seminars, and then at specialist conferences. These provide the precursor to publication in the refereed literature, which may take place before completion of any associated doctoral dissertation. Publication in the professional literature will bring important findings to the attention of interested practitioners.

The research sequence above can easily be squeezed to provide the elements of the traditional positivist approach as devised by House (1970), and illustrated in Figure 2.2. Again this approach assumes the presence of specific conditions:

- the specification of a priori hypotheses: formulated on the basis of theory and literature before any data is collected or fieldwork contemplated;
- the specification of a priori criteria: to measure the acceptability of the hypotheses, most commonly in the form of standard statistical tests;
- the isolation and control of the variables to be investigated: determination of which variable(s) will be treated as dependent, which will be independent (explanatory), and which will be held constant, matched or ignored;
- the verification of the methods for measuring and the variables: specification of which variables can be measured directly, and how, and those which will require the use of proxy variables, or measurement instruments, of some form.

However, we have to acknowledge that there is no single method which necessarily applies to research in all situations. Thus while the positivist tradition remains the most prominent in the accounting literature, non-positivist approaches have become increasingly acceptable. (Even so, some of the top

FIGURE 2.2

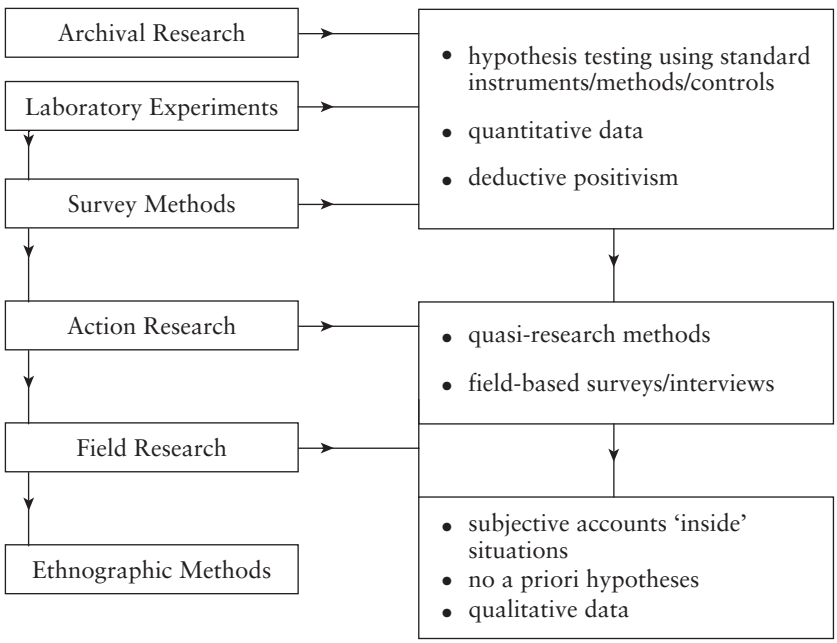


The positivist approach

US journals are still unmoved in their attitude to non-positivist approaches. Baker and Bettner (1997), among others, observe that most of the top journals are devoid of interpretive and critical research studies.) However, management-oriented investigations of change (e.g., the implementation of accounting innovations) may be particularly unsuitable to a scientific approach. Where people are involved and multiple variables are beyond the control of the researchers, including management's own motivation and agenda, scientific approaches are of questionable validity. Checkland (1981, p. 316) observes: 'attempts to apply scientific methodology to real world, essentially social problems, have been responsible for the limited success of management science'.

Thus we can stay within our original 'research sequence' framework but extend out beyond the positivist approach. Figure 2.3 illustrates the range of possibilities. As we move from top to bottom in the figure, we move from the traditional positivist approaches (archival and experimental studies) through field studies towards a case-based approach typically associated with ethnographic studies. This movement corresponds with an increase in the number of uncontrolled variables, with our increasing inability to formulate testable hypotheses, and with the increasing prominence of the 'human' element.

FIGURE 2.3



Alternative research methods

Emergence of the research topic

We should begin by choosing a research topic which is of interest both to the researcher and the supervisor, where the project is contributing to a doctoral qualification. The topic should generate enthusiasm in the researcher at the outset, otherwise he or she is unlikely to last the course of a protracted period of study in which motivation is bound to wane, even temporarily. The source of the topic can be from anywhere, but most commonly:

- a problem at work with potentially wider implications;
- a problem or application spotted in the newspaper or from television;
- a conference presentation revealing the directions being explored by other researchers;
- working papers and completed theses elsewhere – the contents of which are usually at least two years away from publication;
- textbooks – particularly in management-related areas – which are a constant source of untested theories;

- review articles, analysis of the literature in a particular area, to reveal the current boundaries of knowledge and a potential research agenda. The *Journal of Accounting Literature* is a particularly useful source in this regard (e.g., Bauman, 1996; Dunk and Nouri, 1998; Gramling and Stone, 2001; Jones and Shoemaker, 1994). Also good is *Accounting Organizations and Society* (e.g., Hartmann and Moers, 1999; Langfield-Smith, 1997; Libby and Luft, 1993).
- Review monographs (e.g., Ashton and Ashton, 1995, information processing; Brownell, 1995, management accounting; Smith and Gurd, 2000, behavioural issues; Trotman, 1996, auditing) are also helpful, as is Foster's (1986) excellent book, which unfortunately has never been updated since this second edition.
- Refereed journal articles, particularly the final sections, revealing flaws in existing research, gaps in our knowledge and research opportunities.

The ideas will rarely emerge, therefore, from a 'spark' of original thought. More likely, the thought development will have emerged elsewhere, with the originator having either discarded his or her ideas or not seen their full value. It may fit an 'added value' concept in the same way as a successful innovation may be remote from the inventor. This may be in the form of relating two concepts from different disciplines, in a manner which provides an application opportunity in the accounting environment. Here the 'success' is the publishing of research findings in a respectable journal. To do so we will inevitably be building on the work of others.

The common element in each of the above approaches is 'reading' – hence the common advice given to doctoral candidates of 'reading, reading and yet more reading' to know an area and spot the opportunities. Candidates usually have a much greater commitment to a topic if they have developed it themselves, yet many find idea-generation an extremely difficult process. Thus it is not uncommon for the supervisor to be the source of the research idea, because active and experienced researchers usually have far more ideas than they are capable of exploring by themselves. As Gill and Johnson (1997, 2002) observe, topic selection can be risky if left entirely in the hands of the candidate; the chosen topic may prove to be too small, too large, or simply not feasible in the time frame (especially for longitudinal studies). There may need to be a trade-off between the ownership/commitment associated with a student-selected topic and the practicability/timeliness expected of a supervisor-directed preference. Commonly, candidates will contemplate studies which involve the implementation of an accounting change, in order to monitor the change process and the resulting impact on financial outcomes. Such a model is rarely feasible because, apart from the access problems, involvement and data collection will be necessary over a period usually extending beyond that permitted within a standard candidature.

Once the germ of an idea is forthcoming it must be worked over to see if it really constitutes 'research'. For example, are we sure that it is more than a consultancy project? Is it more than a trivial problem with no wider

implications? Is it more than a replication of something someone else has done before, or done in a different industry or different country? If we are happy in this regard then several other questions emerge:

- Is the project 'doable' in a reasonable time frame (e.g., the period of candidature)?
- Will the project fit the NIRD acronym (usually attributed to Rashad Abdel-Khalik during his tenure as editor of *The Accounting Review*) in that it is new, interesting, replicable and defendable? Unsurprisingly, the acronym fits a positivist outlook since replicability may be impossible to guarantee in field study settings.
- Will the data required be readily available? If site visits are required, will access be available over a sufficient time period and to a sufficient depth? This last scenario is of great practical concern and difficult to control. Young and Selto (1993) report on a study whose depth was seriously curtailed because management changed its mind and restricted the access to personnel due to be interviewed. Worse than that, there are numerous cases of change of company ownership during the data collection period, resulting in further site access being permanently denied.

Once the general topic area has been determined, it may be refined by formal methods (e.g., brainstorming, attribute listing, etc.) to identify possible fruitful directions and potentially interesting relationships, and to eliminate blind alleys. Diagrammatic aids, particularly white-boards, are very helpful at this stage for mapping ideas, variables, relationships and processes.

If you are in any doubt as to what constitutes 'research' and what constitutes a study that is published in a reputable refereed journal, then the reader is referred to recent examples from the professional literature. Often the professional journal will have interesting ideas and will convey a message with implications for practitioners (its essential purpose), but will not have the essential conceptual framework of a refereed research piece. Consider, for example, the contribution to the professional literature made by Smith and Briggs (1999). The paper is both amusing and interesting in making observations in a number of areas, and is reproduced in its entirety as Appendix 3. It addresses:

- the traditional stereotype of the accountant as a boring individual with poor interpersonal skills;
- the portrayal of the accountant in the media (particularly in films), initially as a comic incompetent, but latterly as an unethical, unprofessional manipulator with criminal tendencies;
- the accountant, almost exclusively, seen as a white male occupying a subordinate position;
- the concern in the accounting profession about its failure to attract the best candidates, who are lost to other professions;

- the concern that the characteristics which the profession tries to convey about accountants (financial expertise, interpersonal skills, ethical conduct, openness to all) are at odds with those conveyed by the media.

The paper speculates that there might be some link between these issues and suggests that recruitment to the profession is being damaged by the accountant's image. However, it provides no evidence to support this relationship, nor does it provide any evidence for the manner in which this image may be being influenced by the popular media. Furthermore, there is as yet no theoretical explanation to justify these relationships. Without these foundations we have an interesting idea but no basis for a research project whose findings would interest the refereed journals. However, the idea could potentially yield rewarding research projects, were answers to fundamental questions to be found, and these are considered in more detail later in this chapter.

Attention to the fundamental requirements of the refereed literature will allow the researcher to produce an outline research proposal, one that is continually revised during the reading period and may nevertheless have to be revised further during the conduct of the research itself due to unforeseen circumstances. A typical research proposal will include the following elements:

- **Title** – should make it clear what you are trying to do.
- **Abstract** – should summarise the problem, objectives and expected outcomes.
- **Issues** – why they are interesting and important.
- **Objectives** – how the study relates to the problem.
- **Literature** – review of relevant, themed publications.
- **Method** – the how and why of the process.
- **Benefits** – the anticipated outcomes that make this study worthwhile.

A working **title** is important to clarify the topic, especially if external grant income is being sought to support the project. However, the final title is rarely the original one and there are plenty of opportunities to make changes.

The **abstract** is an important departure at this stage, because it allows the researcher to speculate on what the outcomes of the research might be should everything go according to plan. The abstract can be ambitious initially, but will require revision (perhaps radical revision) as problems and constraints emerge in the research process.

The contribution of the paper and the way that it aims to address important **issues** in a systematic manner are fundamental to its success. Internal consistency and overall coherence should ensure that the **objectives** and the intended approach are appropriate.

The outline **literature** review may be incomplete but it must none the less identify the key motivating literatures and theories. Research candidates must recognise that the literature review is a constantly moveable feast and something that will be added to right up to final presentation of the research findings. One of the major deficiencies of both papers and dissertations is

that they frequently overlook the most recent relevant publications: it can be a heart-stopping moment when one is about to submit and the latest issue of a journal appears to report the outcomes of a research project very similar to one's own! At the very least this new paper must be cited. A common complaint from inexperienced researchers is that there is 'no literature' available. If this is true, it may mean that the projected topic may be too trivial to consider. More likely, however, is that the literature review should drill down further and search on different keywords. That there is a dearth of recent literature on a topic may foreshadow problems. For example, papers on 'decision-making heuristics' were common in the late 1970s and early 1980s, and papers on 'group decision-making' common in the mid-1980s, but progress in both of these research areas has slowed, and publications are relatively rare because the psychological theories underlying the research have not developed sufficiently to facilitate new approaches. With respect to apparently new projects, for example, research into 'e-commerce' related topics, students must recognise that e-commerce is just a new way of doing business, and that their review must address the implementation of prior business innovations.

Discussion of **method** should address the alternatives available in order to demonstrate that the preferred choice is the most appropriate. The proposal should also echo the **benefits** of the research, in particular its contribution to knowledge and the potential implications for business practice.

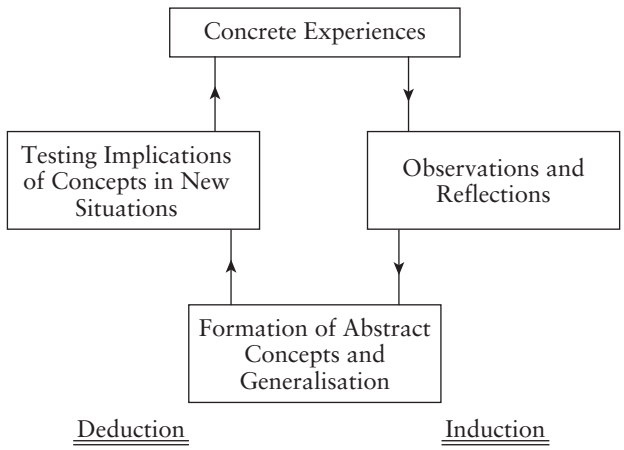
The research proposal would normally form a central feature of any application for ethics approval, and must therefore demonstrate both the value of the research, the integrity of the methods employed and the extent of the involvement of human participants.

Conceptual frameworks

A valuable part of the initial planning process is the development of a conceptual representation of the research project. This can help to clarify the important relationships (and the need for supporting theory), the explanatory and intervening variables, as well as the demonstration of causation.

The inductive and deductive approaches identified in Chapter 1 provide an objective alternative to the conduct of research, but neither allows the opportunity for human interaction: the inductive approach, where new theory is developed on the basis of fresh observations (as is most commonly the case in hard sciences, like astronomy), and the deductive approach, where theory provides the basis for the testing of empirical observations (and which is the most common form of positivist accounting research). The deductive approach is suitable in a highly structured environment, involving the empirical testing of theoretical models, so that its reliability is dependent on the integrity of quantitative and statistical methods. However, the causal relationships

FIGURE 2.4



Kolb's Learning Cycle

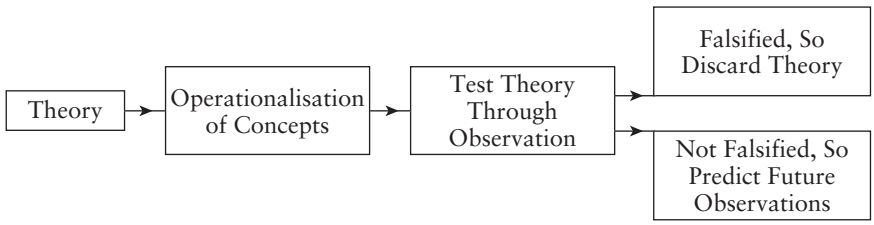
explored rely on an internal logic and take no account of the human relationships present. The application of the inductive approach in the accounting environment necessitates a variation to the traditional model such as that provided by Kolb's Experiential Learning Cycle (Kolb et al., 1979, p. 38), which is illustrated in Figure 2.4. Recognition of the importance of internal processes and human relationships to the inductive approach allows for the existence of human subjectivity without distorting research findings, even though they may be qualitative and not replicable. Where human relationships are central to an understanding of accounting behaviour the approach exploits the subjective environment.

Although both models provide opportunities in accounting research, the deductive approach offers greater possibilities for the implementation of scientific methods, since it facilitates arguably more reliable measurement and control. Grounded theory (discussed in more detail in Chapter 8) illustrates the potential for inductive methods in accounting research.

We can therefore develop the model of the deductive process (see Figure 2.5) so that it corresponds with Popper's (1959) defining characteristics of scientific theory:

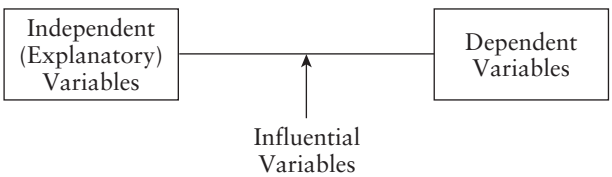
- the theory is capable of empirical testing;
- scientists (and researchers) should make rigorous attempts at falsifying theory;
- science advances as falsified propositions are left behind, leaving a core of theory still to be disproved.

FIGURE 2.5



The deductive process

FIGURE 2.6



The conceptual schema

The basic conceptual schema (see Figure 2.6) provides a powerful tool for the examination of causal relationships in a positivist environment. By establishing the key variables of interest, and the other potentially influential factors, we can form a better impression of the breadth of the problem.

If, for example, we return to our earlier article on 'accounting stereotypes', we can see how the content of a professional piece may be used to develop a conceptual schema for more systematic investigation. As a starting point we need to identify dependent variable(s) – one or more outcome variables whose value may be influenced by a number of (independent) explanatory factors.

The scenario outlined extends over a long period of time – potentially from the final years of secondary school through undergraduate years to recruitment to the accounting profession; quite possibly in excess of five years even for straightforward cases. The length of this period, the process of maturation and the incidence of a multitude of uncontrolled variables makes it extremely unlikely that we will be able to predict behaviour in year five based on attributes in year zero. In trying to establish a causal relationship for investigation, it is much more realistic, therefore, to break up the process in

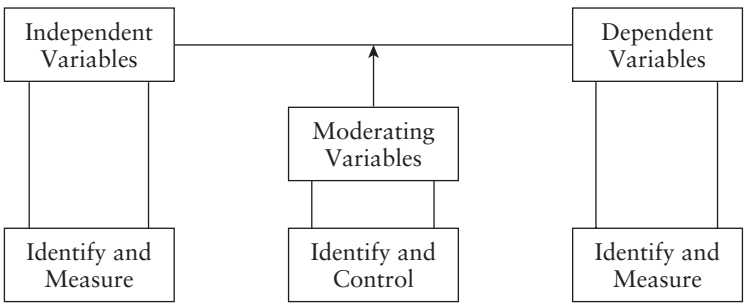
order to be able to focus on events during more finite time periods. Such a reorientation allows us to focus on two 'choices' which will permit the suggestion of two measurable dependent variables:

- the selection of commerce/accounting as the preferred undergraduate course by school leavers; and
- the decision to enter the accounting profession by those nearing the end of their university course.

The major dependent variable, in each case, as suggested by the professional article, is the 'image' of the accountant, though it is unclear how this concept is to be either defined or measured. In each case there is likely to be a potentially large list of influential variables that may impose conditional relationships: for the former, these include teachers, parents, schoolmates, perceptions of other courses, job prospects, etc.; and for the latter, these include income, career prospects, perceptions of competing professions, etc. Both of these 'models' in their simplest form ignore how the perceptions of the accounting profession have been formed (e.g., did film and media have any influence whatsoever, or are there other more important influences?), and whether entry into the profession can be regarded as a single decision, irrespective of the claims of the competing recruiters (e.g., individual members of the Big 5, the public sector, commerce).

In terms of our example(s), we can only sample relatively small proportions of both school-leavers and potential recruits to the profession. In neither case will we have access to population statistics. In addition, there may be problems with interviewing school-leavers prior to university entrance, especially if we are contemplating the conduct of psychometric testing to examine personality issues and different cognitive styles. Such an investigation is better conducted when the decision has been made, early in the first year of university courses. This has the advantage of providing access to mature entrants and of looking at an actual decision rather than a choice intention, but it has the disadvantage of potential hindsight bias. Data collection at this time may determine the motives for choice of entry into competing professional courses (e.g., accounting, law, medicine). For the decision point prior to entry into the accounting profession, some collaboration with the major recruiters (e.g., the Big 5, state government) is advisable; they are all interested in attracting the 'best' students and make clear statements on what they are looking for among successful applicants. The researcher may gain access to student information, with appropriate permissions and participation, to identify those selected for interview and those subsequently offered positions. A comparison of recruiter decisions with student profiles should facilitate an examination of the transparency of recruitment policies. The design problems associated with the identification and measurement of variables means that the basic schema can be modified (see Figure 2.7).

FIGURE 2.7



Measurement issues

In terms of our example, we have serious problems in defining and measuring the dependent and independent variables, and in controlling influential variables:

- 'Image' cannot be measured directly. The literature will help in identifying those variables which contribute to the creation of the concept. This process may allow the identification of variable(s) that can be measured directly and used as a proxy for 'image'. This is a far from satisfactory arrangement but may be the best we can do if the investigation of these relationships is to proceed.
- Decision choices will have been influenced by many variables over long periods of time. We must restrict our focus to those variables cited in the literature as being potentially influential, and try to gauge their relative importance to decision-making in this context, possibly through pilot surveys or interviews after the event. The use of matching procedures (e.g., sampling candidates from the same schools, of the same age and gender, and with equivalent academic profiles at entry) will help to provide some control over the influence of extraneous variables.

We clearly have some way to go before this interesting research idea becomes a feasible research project. Questions relating to theory and measurement remain problematic, but a search of the literature specifies a number of theories with precise implications for this scenario. Friedman and Lyne (2001) examine the development of the 'beancounter' stereotype and its implications for the future of the accounting profession. They identify a number of sociological theories which could assist the establishment of our conceptual framework:

- Hamilton and Troiler (1986, p. 13) suggest that stereotyping is determined by cognitive, motivational and social learning processes. *Social*

learning theory suggests that stereotypes are mostly acquired through socialisation, via variables such as parents, schools and mass media, but that they may be modified by subsequent experience.
- Campbell (1965) suggests that *realistic conflict theory* may explain how stereotypes result from inter-group competition for scarce physical resources.
- Tajfel and Turner (1985) develop a *social identity theory* which modifies conflict theory to take account of individual achievement motives so that conflict can also occur over scarce social resources like prestige and status.
- Stroebe and Insko (1989, p. 14) recommend the simultaneous consideration of the three approaches: social learning theory, social identity theory and conflict theory.

The implications of these theoretical approaches to our research idea are instructive:

- The perception of a negative accounting stereotype is unlikely to fade among school students or accounting undergraduates while it continues to be widely perpetuated (social learning theory).
- The negative associations are likely to impact on candidates' choice of career profession and the associated income and prestige (social identity theory).
- The inter-group conflict between the professions, when each is trying to attract the 'best and brightest' candidates, means that they may seek to perpetuate the negative stereotypes of other professions to their own resource advantage (realistic conflict theory). The inference here is that we may extend the research to consider the advertising material used by the competing professions (and by competitors within the accounting profession) and exposure of students to these. We may anticipate that recruitment at school level involves the denigration of the accounting profession by other professions, and at university level the denigration of some members of the profession by fellow members from rival organisations.

What is also apparent is that the simple conceptual schema of Figure 2.6 has outlived its usefulness in this context. It was helpful to begin with but must now be modified to allow the further development of hypotheses and data collection methods. Chapter 3 focuses on these issues in the development of the important links between theory, literature, hypotheses and methods.

In Chapter 1 we acknowledged our debt to researchers in the natural and social sciences. It is now helpful to turn to popular literature, which provides highly readable explanations of complex situations, for an insight into the emergence of research ideas, the development of research questions and the airing of potential solutions for testing, to illustrate the research sequence in practice. It is instructive to consider the different scope of original research in solving practical problems by examining the particular aspects of three well-documented stories from non-accounting environments. Thus we explore the

development of new theory in a chemical environment in 'The Structure of DNA', the development and testing of alternative strategies to address a sporting issue in 'The Bradman Problem', and the solution of apparently insuperable implementation issues to a problem where the 'answer' was well known in 'The Longitude Problem'.

The structure of DNA: the development of new theory

James D. Watson's (1968) *The Double Helix* (subsequently filmed as *Life Story*) provides a brilliant description of the exciting process of discovery in scientific research, even if the approach adopted is rather unorthodox. The development of theory and conceptual modelling, from systematic deductions based on the empirical findings of others, is conducted in a competitive environment where the ultimate prize for winning the 'race' is the Nobel Prize. (Watson, together with Francis Crick, both of Cambridge University, and Maurice Wilkins, of King's College, London, were awarded the 1962 Nobel Prize for Physiology of Medicine for their pioneering work during 1951–52.)

Sir Lawrence Bragg reflects on these achievements by researchers in his Cavendish Laboratory through a revealing preface to Watson's book, with implications for research ethics:

> He knows that a colleague has been working for years on a problem and has accumulated a mass of hard-won evidence, which has not yet been published because it is anticipated that success is just around the corner. He has seen this evidence and has good reason to believe that a method of attack which he can envisage, perhaps merely a new point of view, will lead straight to the solution. An offer of collaboration at this stage might well be regarded as trespass. Should he go ahead on his own? (Bragg, in Watson, 1968, p. vii)

That Watson and Crick, at the Cavendish, did proceed with a belated relationship – though collaboration is too strong a word – with Wilkins and Rosalind Franklin is a matter of history. The course of their investigations, and the factors leading to their success in developing a new theoretical model, have implications for all research. Essentially, they proceed to develop a model which fits all of the evidence currently available to them, and await confirmation or disconfirmation of their framework from the empirical findings of others:

- The work of Linus Pauling in the USA on a helix formation for polypeptide chains suggested that DNA (deoxyribose nucleic acid) too had a helical structure. Pauling's early attempts at modelling though, without crystallographic evidence, had produced stereochemically impossible components. Early evidence from X-ray crystallographic diffraction presented by Wilkins also seemed to suggest a helical structure, but there was no

evidence of whether a single, double or triple strand helical configuration was most appropriate. Crick and Watson apparently proceeded on the basis of an educated guess favouring the double helix because most things biological come in twos!

- Ernst Chargaff had produced vital evidence on the ratios of constituent bases, and particularly the equalities existing between adenine (A) = thymine (T), and guanine (G) = cytosine (C). The A–T, G–C flat hydrogen-bonded base pairs formed the core of the Crick and Watson structure, rather like a spiral staircase in which the bases form the steps.
- Important advice from a structural chemist colleague (who just happened to be sharing the same office with Crick and Watson) suggested that the normal textbook formulation of the A–T, G–C bases was incorrect and that they should work with an alternative 'keto' form. Without this important questioning of textbook content, and accepted knowledge, their structure would not have held together. The impact of this finding was that a given chain could contain both purines and pyrimidines (with the capacity to carry the genetic material for self-replication) and that the backbones of the chains should run in opposite directions.

Thus Crick and Watson were able to construct a physical model comprising two intertwining helically coiled chains of nucleotides, right-handed and running in opposite directions, with complementary sequences of hydrogen-bonded bases. The resulting structure was stereochemically possible, and subsequent X-ray evidence from Franklin confirmed that the sugar-phosphate backbone was indeed on the outside of the molecule.

The more general implications for researchers are constant vigilance and a questioning attitude to the work of others and existing publications.

The Bradman problem: the development of new strategies

Test match cricket in the 1930s was dominated by a single, outstanding individual whose unique gifts of batsmanship threatened to change the way the game was played. Donald Bradman scored so many runs, and scored them so quickly, that he was likely to win a game almost single-handedly. Ashes test series, once so closely contested, threatened to be one-sided affairs, with Australia the perennial victor. In the 1930 England v. Australia test matches Bradman totalled a record 974 runs at an average of 139.14 per innings, and recorded separate scores of 334, 254 and 232. At the commencement of the 1932/33 series he averaged 112.29 in all test matches, and had already posted six scores in excess of 200 in only 24 completed innings.

Bradman was a phenomenon, some would say a 'freak', and curtailing his dominance became a pressing question for successive England captains.

'The Bradman Problem' is capably detailed by Lawrence Le Quesne in his book *The Bodyline Controversy* (1983), and presents an intriguing research question, investigating a number of alternatives that might provide successful solutions.

1. Changing the playing conditions

Cricket was played on hard, fast wickets, largely true, though with the occasional unpredictable bounce. If bowlers did not take wickets in the first few overs, when the ball was still shiny and swinging, then they might have a lot of overs to bowl before they got a replacement ball, by which time 200 runs had been scored. Wickets were uncovered and exposed to the elements once a game had started and could become unplayable as a hot sun dried out a wet pitch; these 'sticky' wickets provided a possible solution to the problem because Bradman was nowhere near as prolific on bad wickets as many of his contemporaries were. However, captains could hardly rely on this occurrence to blunt Bradman's genius on a regular basis.

2. Changing the rules of the game

Test matches were timeless in the 1930s and played to a finish. Declarations were rare and slow play very common. Further, the leg-before-wicket (LBW) rule made it difficult for the batsman to get out in that way – he had to be caught in front of the stumps by a ball pitching in line, wicket-to-wicket. Life was difficult for bowlers, and a number of other batsmen (notably Ponsford for Australia and Hammond for England) regularly completed double and triple centuries, although none with the regularity, reliability or speed of Bradman.

Reduction of the specified playing time to five (or four) days would mitigate slow play and dull the impact of less talented batsmen content to occupy the crease, but would not affect Bradman – witness his 309 in a day against England at Leeds in 1930. However, timeless tests were outlawed before the end of the inter-war period. Similarly, the LBW law was changed so that batsmen could be given out to a ball breaking back from outside the off stump and striking the pads in front of the wicket.

3. Changing the bowling

Bowlers had often sought to restrict the batsman by bowling outside of the leg stump (e.g., Hirst and Foster in the early 1900s), ostensibly to cut out off-side shots but also to restrict the on-drive through bowling just short of a length. An accurate bowler could therefore depress the scoring rate by keeping the batsman to deflected singles in the arc between the wicketkeeper and square leg. As a consequence the whole game is slowed to a snail's pace – a

strategy still used today to curb an aggressive batsman's dominance. But Bradman was too good to fall for such tricks and used his agility, quick reflexes and nimble footwork to move to leg, outside the line of the ball, so giving himself room to hit through the off-side field. Such a strategy would have made less gifted batsmen vulnerable because it entailed the exposure of all three stumps to the bowler, but Bradman could execute the manoeuvre without undue risk.

4. Changing the field placings

An orthodox field will involve the placing of fieldsmen on both sides of the wicket, with a majority coincident with the bowler's planned line of attack. This means plenty of gaps in the field and, for inaccurate bowlers, plenty of scoring opportunities. Clearly, a preferred strategy would be to place all the fielders in a tightly confined space and have the bowler deliver a line and length that forces the batsman to play the ball there. This is a sound strategy for an accurate bowler – witness Laker's devastating use of a packed leg-trap to a turning ball in 1956 – but for a batsman of Bradman's ingenuity we now have a vacant off-side to be penetrated by inventive, unorthodox shots.

5. Introducing a 'bodyline' attack

While a combination of numbers (3) and (4) above provides a partial solution in restricting scoring opportunities, only impatience will induce eventual dismissal. For Bradman it may do neither. However, together they provide the basis for a potentially successful solution: 'leg theory' will dictate the line of delivery, but we need also to control the height of the delivery to induce the batsman to give chances from playing the ball in the air. Herein lies the Jardine–Larwood proposition, initiated by Douglas Jardine, England's captain in 1932 and executed brilliantly by Harold Larwood, the fastest and most accurate bowler of his time.

Extreme fast bowling does not provide even Bradman with the time to move outside the line consistently without risk. The introduction of a high proportion of short-pitched balls rearing towards the throat or the rib-cage of the batsman makes scoring without risk extremely difficult. Batsmen are likely to fend off the ball defensively – to be caught in the leg-trap ring of close catchers. If they try to attack by hooking the ball over the in-field, they fall prey to a number of deep-set fielders on the leg-side and behind the wicket. Concentrating almost all the fielders in the arc between wicketkeeper and mid-wicket, five close to the bat and three close to the boundary, covers almost all the options. Scoring is restricted to risky options and bad balls.

This form of attack was 'successful' in that it resulted in a 4–1 series victory for England and also provided an appropriate solution to 'the Bradman problem' in that he scored only one century in the series, a total of only 396 runs at an average of just 56.57. He still tried to play in a cavalier fashion,

moving outside the line to play tennis-like shots; the result was brilliant, hectic cameos that were over all too soon for his team's requirements.

However, the risks of bodily injury to the batsman from the bodyline solution were high, and its introduction was seen to be ungentlemanly and against the spirit of the game. Jardine and Larwood were never chosen to play against Australia again. The subsequent 1934 test series was very much a fence-building exercise, with the England bowling friendly and 'bouncer' free. Bradman was again unconstrained and scored 758 runs at an average of 94.75. The success of the leg theory solution generated further changes to the rules of the game, with a restriction on the number of fielders permitted on the leg-side behind the wicket and the number of short-pitched balls that were allowed to be bowled per over.

The more general implications for research are that there may be legal, moral, ethical or professional circumstances which prevent either the conduct of the research or the implementation of recommendations from the research findings.

The longitude problem: implementing solutions

The measurement of longitude at sea requires the accurate measurement of both the time at the current location and that at the Greenwich meridian, or some other similar base point. The time difference allows the calculation of geographical separation – since the 24-hour revolution of the Earth constitutes a 360 degree spin – so that a one-hour time difference constitutes 15 degrees of longitude, where one degree of longitude is equivalent to 68 miles on the Equator. The measurement of local time is not a problem, especially during hours of daylight, but in the absence of accurate timepieces knowledge of the corresponding time at the base point remains a considerable problem.

The consequences of being unable to measure longitude were serious and have been detailed by Dava Sobel in her book *Longitude* (1995). Shipwrecks and lost vessels were common, and piracy was facilitated by the need for ships to track across common lines of latitude on the 'trade routes' to maintain their position. The pendulum clocks of the 1660s, due to Huygens, had been used to demonstrate the possibility of measuring longitude at sea with timepieces, but they were only helpful in favourable weather. So much so that Sir Isaac Newton (cited in Sobel, 1995, p. 52) was forced to admit in 1714: 'One [method] is by a watch to keep the time exactly. But by reason of the motion of the ship, the variation of heat and cold, wet and dry, and the difference of gravity in different latitudes, such a watch hath not been made.' Newton clearly had astronomical, or at least scientific, solutions in mind rather than mechanical ones, necessitating the consideration of alternative solutions.

1. Existing methods

These were largely confined to 'dead reckoning' and 'compass method' approaches. Dead reckoning required estimates to be made of the speed of the ship, in conjunction with calculation of the effects of wind speed and currents. Its success relied on good seamanship, reliable maps and luck! Compass methods were concerned with comparisons between magnetic north and 'true' north as shown by the pole star. Relative positions allowed the estimation of longitude without the necessity of measuring time. However, compass needles were notoriously unreliable, with a great deal of variation for the same compass on successive voyages. This, coupled with variations in terrestrial magnetism, made readings highly dependent on the particular seas being traversed.

2. Eclipse data

Eclipse data were thought to be potentially useful. Sonar and lunar eclipses provided opportunities if it was known when they were expected to be observed in other locations, but such occurrences were far too rare to provide a realistic navigational aid. Galileo had established that eclipses of the moons of Jupiter were extremely common and predictable, making them an accurate means of specifying longitude at specific land-based locations. However, movement aboard ship made this an impossible strategy for navigation, even when the night skies were clear.

3. Lunar distances methods

These involved measuring the distance between the moon and the sun, by day, and between the moon and stars at night. Such methods required detailed data on the track of the moon and the positions of individual stars so that the disappearance of particular stars behind the moon could be predicted. The complexities of the moon's orbit impeded progress in the prediction of the required measurements at different locations, and it was not until 1725 that Flamsteed's posthumous almanac of star positions was published. Even so, the available tables still meant it took four hours to calculate longitude (subsequently reduced to 30 minutes by Maskelyne's 1766 almanac).

The lunar distance method was therefore shown to be a theoretically possible means of accurately computing longitude, made more practicable by the invention of the quadrant (later sextant) in 1731 to measure elevations of, and distance between, moon and sun by day, and moon and stars by night. This permitted an estimate of time differences between a ship and known, fixed land locations. Even so, actual measurement proved impossible at times for a variety of reasons:

- weather conditions occasioning fog or thick cloud cover;
- the moon is so close to the sun for about six days per month that it disappears from view; and
- the moon is on the opposite side of the Earth from the sun for about 13 days per month.

John Harrison adopted a more direct solution to the problem, questioning the position of Newton and proceeding to build a succession of clocks that were shown to be accurate to fractions of a second per day. By eliminating problems of friction, he developed clocks that required no lubrication or cleaning. This, combined with the use of bi-metal strips of non-corrosive materials, overcame the problems of temperature change and rust. The choice of innovative balancing mechanisms also meant that the clocks were virtually unaffected by the most severe weather conditions.

By the time Harrison died in 1776 copies of his watch were still rare and expensive (in excess of £200) whereas a good sextant and set of lunar tables could be purchased for as little as £20. This considerable price difference meant that the 'lunar distance method' of calculation remained prominent until more affordable watches became available in the early 1800s through the sale of the mass-produced Arnold/Earnshaw pocket 'chronometers'.

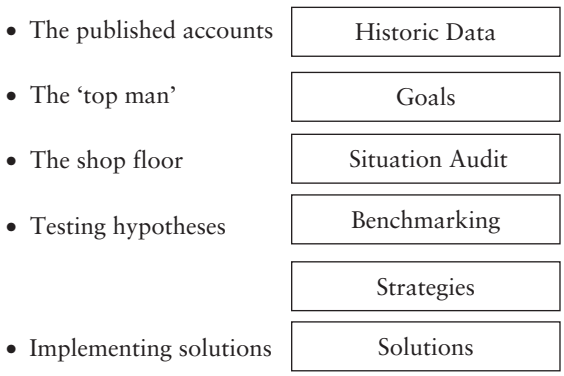
Both timeliness and resource cost remain fundamental elements in the conduct of research projects and the implementation of their findings.

The scope of these three examples is very different, concerned respectively with the development of new theory, the development of workable solutions and the implementation of workable solutions. As we suspected, the research process is neither simple, systematic nor clean in any of the cases. What is common throughout are the pivotal roles played by 'theory' and 'validity': good theory produces good findings, and we are able to evaluate both the reliability and the validity of these findings through external reference. The following chapter examines theory in more detail, and expands the consideration of reliability and validity as desirable characteristics of accounting research.

Strategic management accounting

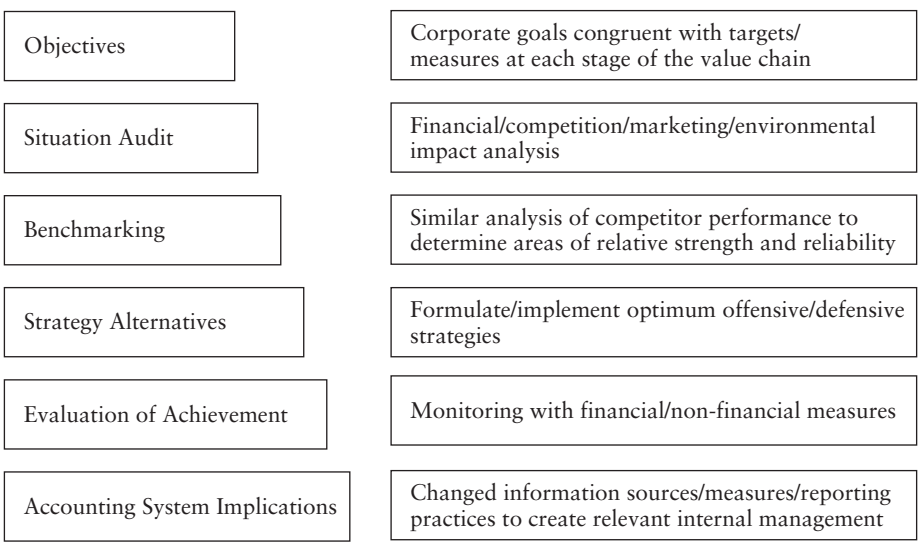
Within an accounting environment we can observe the identification, definition and solution to practical business problems, solutions which require the careful specification of the research question, the development of hypotheses and alternative implementation strategies. Thus John Harvey-Jones (1992; Harvey-Jones and Mossey, 1990), the celebrated 'troubleshooter' of the eponymous television series, details his approach to the investigation of practical issues in real business situations. His activities might be considered to correspond

FIGURE 2.8



The Harvey-Jones approach to problem-solving

FIGURE 2.9



Generalised process–improvement sequence

to a form of action research in that he is actively collaborating onsite with other individuals, although his practices (especially those apparent in the *Morgan Cars* episode) rarely correspond with accepted 'good practice' in action

research. Harvey-Jones appears to develop a systematic model during the first series for application in the second series (*Troubleshooter* 2). Central to this model is the specification of a fundamental research question based on an analysis of the published financial accounts and empirical observation of the site. Interestingly, this specification of the research question rarely coincides with that of the CEO of the organisation concerned. Further observation and benchmarking against the performance of other organisations allows the development of hypotheses and alternative approaches that may yield the desired outcome. Reporting of the recommendations often causes conflict with the senior management of participating organisations, especially in the earlier episodes, where the importance of organisation culture to an acceptable solution appears to have been underestimated or overlooked. The stages of analysis depicted in Figure 2.8 generate the more generalised framework for business solution and improvement opportunities in Figure 2.9.

It is apparent that we have here a systematic, almost strategic, approach that owes relatively little to theory. An accepted positivist research sequence offers great similarities, except that a much wider preparatory stage is undertaken, which is rarely restricted to the consideration of a single case. Specification of the individual elements of the research sequence here provides the basis for their detailed discussion in Chapter 3, with the focus on theory and literature sufficient to be able to build testable hypotheses.

3

Theory, Literature and Hypotheses

CHAPTER CONTENTS

We have established theory, reliability and validity as three of the characteristics of accounting research that we wish to achieve. In this chapter we expand this list to five: good theory, reliability, construct validity, internal validity and external validity. We also recognise that because of the inherent trade-offs we can only have some part of each of these characteristics at the same time; choices and compromises are necessary. In this chapter we continue to explore these aspects of good research so that we are in a position, in Chapter 4, to choose between alternative methods in identifying the most appropriate approach to a particular research question.

Sources of theory

We begin with some definitions to overcome potential confusion with terminology, especially where the edges to some concepts appear blurred and the definitions interdependent.

- **Theory** is a network of hypotheses or an all-embracing notion that underpins one or more hypotheses. In Chapter 1 we described theory as 'a set of tentative explanations' with which to justify diverse observations. We

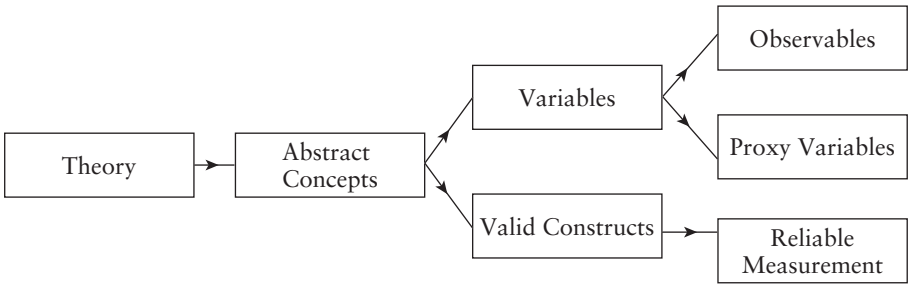
need a theory to have some justification for expecting a relationship to exist. Where we have none our *hypotheses* are immediately disputable.

- **Hypotheses** are supposed relationships, possibly causal links, between two or more concepts or variables. A hypothesis should be testable, but it may not be directly so if it comprises a number of abstract *concepts*.
- **Concepts** are abstract ideas, not directly observable or measurable, which must first be 'operationalised' in some way to provide measurable indicators. This will be achieved either by identifying a variable that is an adequate substitute for the concept or by developing a *construct* to provide a new measure of the concept.
- **Constructs** are indirect measures of concepts usually generated in the form of multi-item questions. The sum of a set of valid and reliable responses to the construct provides a measure of the concept.
- **Variables** are observable items which can assume different values. These values can be measured either directly or, if this cannot be done satisfactorily, indirectly through the use of proxy (substitute) variables. Variables are usually independent (i.e., explanatory), dependent (i.e., are explained by the independent variables), moderating (i.e., have a conditional influence) or intervening (i.e., with an influence, potentially spurious, that needs to be controlled).
- **Reliability** establishes the consistency of a research instrument in that the results it achieves should be similar in similar circumstances. Thus, the same survey subjects (participants) using the same instrument should generate the same results under identical conditions.
- **Validity** measures the degree to which our research achieves what it sets out to do. We would usually distinguish between construct validity, internal validity and external validity, each of which will be addressed later in this chapter and in subsequent chapters.

We can illustrate these terms with reference to particular examples in the accounting literature, and many further examples will arise in subsequent chapters. The source of most **theory** in accounting research comes not from the accounting literature but from the economics (and finance), behavioural and sociology literatures. Thus Smith and Taffler (2000) use signalling theory to examine the nature of corporate disclosures, in the expectation that firms will behave in a manner that 'signals' to the market that they are high achievers and are adopting industry best-practice. They use this as a basis to establish a formal **hypothesis**, for subsequent testing, that the positive content of corporate narratives will be directly associated with the financial performance of the company.

Brownell (1982) establishes a **concept**, termed 'budgetary participation' in his study, as one of the desirable attributes of leadership in management. He is unable to measure this concept directly so he uses the Milani (1975) **construct** to operationalise it instead. This multi-item instrument measures influence, involvement and participation in budget-setting but is deemed a satisfactory indicator of budgetary participation.

FIGURE 3.1



Searching for construct validity

Much of the literature examining agency relationships (e.g., Watts and Zimmerman, 1978) uses 'political influence' as one of the variables under examination, but because of the difficulty in observing and measuring this **variable** directly, company size (measured by assets or employees) is often used as a proxy variable.

The formal **reliability** of the research instruments employed is rarely addressed in the accounting literature; only survey-based research makes a virtue out of the use of measures like the Cronbach alpha to detail the degree of confidence we have in the means of data collection. Questions of validity are best considered in the trade-offs we have to make, usually between reliability and **construct validity** on the one hand, and between **internal validity** and **external validity** on the other. Thus, in the example cited above, Brownell chooses to use the well-accepted Milani instrument. Reliability is not in doubt because this instrument consistently generates high Cronbach alphas, but construct validity is in doubt since Brownell wants to measure 'participation' in budget-setting but the chosen instrument also measures 'influence' and 'involvement'. The construct being used may therefore not be measuring exactly what is required or being specified. Figure 3.1 illustrates the problem we encounter in striving to achieve construct validity.

We start with a theory which establishes a series of relationships between ill-defined concepts. These may be difficult to pin down and impossible to measure directly. We first search for an observable variable which may act as a proxy for the concept. For example, 'absenteeism' is often used in the management literature to proxy for the concepts 'morale' and 'team spirit'. But if we consider these to be inadequate proxies, then we need to search for an alternative, perhaps by identifying a construct that satisfactorily measures the concept. Ideally, an established construct will already be in existence to measure exactly what we want – this is the best of both worlds: reliability and construct validity. More often we are faced with a dilemma: either use an established construct which does not quite hit the target (threatening construct validity) or develop a new or adapted instrument, by devising a revised set of

questions, which does hit the target (threatening reliability). The former trade-off is the one most likely to be encountered in the accounting literature, though we might argue that we would prefer to see more of the latter.

If we have **internal validity** then we are able to eliminate rival hypotheses with confidence because we can specify causal relationships; we know what is causing what because we are controlling for all other influential factors. This scenario only precisely fits experiments under laboratory conditions, conducted under strict control and perhaps based on unrealistic assumptions. The findings may have no **external validity** whatsoever; they cannot be generalised to the 'real world' because they only apply in the laboratory. This is another fundamental trade-off, and one where we may have to compromise loss of internal validity (loosening the confidence we have in the relationships) in order to increase external validity (and realism).

Both construct validity and internal validity are wholly dependent on good theory: establishing a research design with appropriate concepts, which are underpinned by theory and which are realistically linked to their means of measurement. It is instructive to turn first to the sources of theory available to accounting researchers and to look at these in more detail.

The fundamental distinction underpinning accounting theory is that between normative theory (of what ought to be) and positive theory (of what is or will be). Much of the pioneering work in financial accounting (e.g., Littleton, 1933; Paton and Littleton, 1940) was an embodiment of current practice to establish what *ought* to be the optimal accounting practices, particularly for income determination purposes. Such research offered little opportunity for empirical testing or for the development of positive theory, concerning, for example, how managers and investors would actually react to the provision of new accounting information in their decision-making. We look particularly at developments in economics, finance and organisational behaviour for sources of testable theory that may be applied in an accounting environment; the coverage here is necessarily not comprehensive, but illustrative of the potential sources in other disciplines.

Economics

Early researchers (e.g., Canning, 1929; Edwards, 1938) used economic analysis in a deductive manner to develop alternative approaches to income determination, theorising which was subsequently combined with normative findings by Bedford (1965) to provide a conceptual framework for income determination. Much earlier still, the neoclassical theory of the firm had established an economic framework with fundamental assumptions: decision-making by rational, profit-maximising individuals, working under conditions of certainty, and with freely available information. These basics have had far-reaching effects, leading, for example, to the development of normative decision-making models of practices which would yield profit-maximising outcomes under appropriate assumptions (e.g., those associated with

linear programming solutions, cost–volume–profit analysis and discounted cash flows). But neoclassical theory is unhelpful in providing guidance on the practical behaviour of individuals without significant modifications to the general theory, and a relaxation of its key assumptions:

- Friedman (1953) established the positivist economics perspective: that the purpose of theory is to enable us to make verifiable predictions. In doing so he suggested that even if theory makes unrealistic assumptions, it does not matter so long as verifiable predictions result. These sentiments provided the impetus for the positivist approach to accounting typified by Watts and Zimmerman (1986).
- Simon (1959) developed the concept of 'bounded rationality', which permitted the emergence of 'satisficing' rather than 'maximising' behaviour.
- Kahnemann and Tversky (1972) recognised that decision-making under conditions of uncertainty required the development of appropriate behavioural theories, theories which have lead to a stream of accounting research concerned with decision-making heuristics and decision processes (e.g., Joyce and Biddle, 1981; Smith, 1993).
- Demski and Feltham (1976) explored information economics theories.
- Watts and Zimmerman (1978) developed the concept of managerial self-interest (itself a neoclassical notion) as part of a principal–agent relationship, forming what they termed 'a nexus of contracts' between managers and shareholders, and between managers and subordinates.
- Williamson (1979) developed a theory of economic organisation to explain why activities are organised in particular ways and how choices are made, with implications for accounting research on decision processes (e.g., Spicer and Ballew, 1983).
- MacIntosh (1994) details the labour process paradigm devised by Marx and Engels and which was based in the tradition of political economics. This view places the manager in the position of both victim and user of management accounting information and control systems, and provides for a stream of accounting research in the radical structuralist tradition (e.g., Tinker, 1980).

Each of these developments has had a wide influence on accounting research, particularly in terms of decision-making processes, the motives of the decision-maker and the way in which accounting information is used. They have coincided with the prominence of the decision usefulness approach to financial accounting, especially since the 1970s, prompting empirical developments in the pursuit of a conceptual framework for accounting – a body of knowledge underpinning the discipline.

Finance

Although we might perceive finance as a sub-discipline of economics, the developments in this field have had such a radical influence on accounting

research that they deserve separate consideration. Advances in finance theory have had implications particularly for research in financial management, corporate policy and investor behaviour:

- Markowitz (1952), with work on portfolio risk, lead to Sharpe (1964) and the capital asset pricing model. This formed the basis for the pioneering work of Ball and Brown (1968), linking stock market reaction to the provision of accounting information.
- Modigliani and Miller (1958) developed a theory regarding the risks associated with capital structure.
- Fama (1970) developed the notions of market efficiency associated with the processing of stock price information.
- Black and Scholes (1973) formulated the option pricing model as a vehicle for handling decision-making under uncertain conditions.
- Jensen and Meckling (1976) first expounded the agency costs argument associated with debt–equity trade-offs, which initiated a stream of research linked to the choice of accounting policy, and subsequently to management accounting (see Baiman, 1982).
- Ross (1977), following Spence (1973), initiated incentive-signalling theories in finance, spawning a research stream concerned with voluntary disclosures in financial reporting.
- Scott (1981), following Myers (1977) and the financial economics tradition, developed a theory of corporate failure based on cash flows, underpinning some of the earlier work in failure prediction (e.g., Altman, 1968; Beaver, 1966; Taffler, 1983).

Organisational behaviour

Robbins (1995) observes three levels of research interest in organisational behaviour, associated respectively with individuals, groups and organisation systems, and he develops a theoretical framework for each. Models at the individual level are heavily influenced by the psychology discipline, while those at the group and organisation level are influenced more by sociology and social psychology. All have been highly influential in the development of accounting research.

The **individual** focus looks at contributions to knowledge from the impact of, for example, learning, motivation, personality, perception, training, leadership, job satisfaction, decision-making, performance appraisal, attitudes and behaviour, together with their relationship to biographical characteristics. A selection of the multitude of theories generated in this area illustrates their potential accounting applications:

- Kelley (1972) developed *attribution theory* to explain how we perceive people differently depending on the meaning that we attribute to a given behaviour. This would have relevance to management accounting research based in the appraisal of subordinate performance by managers (e.g., Mia, 1989).

- Bandura (1977) developed *social learning theory*, suggesting that we learn through both observation and direct experience so that individual perceptions can be influenced by teachers, peers and the media. This would impact on the development and maintenance of accounting stereotypes and impact on accounting recruitment (e.g., Friedman and Lyne, 2001; Smith and Briggs, 1999).
- Festinger (1957) developed *cognitive dissonance theory* to explain the relationships between attitudes and behaviour, and the potential impact of conflict therein on both individual and organisation. The theory is relevant to studies of accounting research involving conflicting information or where messages conveyed are alien to the user (e.g., Smith, 1998b; Smith and Taffler, 1995).
- Bem (1972) developed *self-perception theory* to argue that attitudes are used to rationalise behaviour choices after the event. Such outcomes are evident in the incidence of hindsight bias in interview-based accounting research.
- Early motivation theorists (e.g., Herzberg, 1966; McClelland, 1967; Maslow, 1954) identified various intrinsic and extrinsic rewards which motivate performance. More recently, *expectancy theory* (Lawler, 1973) has become the most widely accepted explanation of motivation, arguing that behaviour will depend on the likelihood of our attaining an attractive reward. This theory has been used in the accounting literature by Ronen and Livingstone (1974) to suggest that subordinates will only expend effort in the expectation that their actions will provide intrinsic and extrinsic satisfaction. It has subsequently been used to explore the major behavioural variables linked to the motivation effects of budgets (e.g., Ferris, 1977; Rockness, 1977).

The **group** focus looks at the contribution to knowledge of the impact of group dynamics and processes, communication, behavioural and attitude changes, norms, roles, status, power and conflict:

- Argyris (1952) and Becker and Green (1962) use contingency theories to explore the impact of group dynamics on the budgetary process.
- Barrow (1977) developed a contingency theory of leadership in trying to explain successful leadership and consequent group behaviour as a combination of specific leadership styles and situational conditions.
- Vroom and Yetton (1973) developed the *leader–participation* model, a popular contingency variant, to relate leadership behaviour to participation in decision-making, while emphasising the importance of task structure. This theory has clear implications for accounting research in areas such as budget-setting.

The **organisation systems** focus looks at the contribution to knowledge of the impact of organisation culture and change, structure and hierarchy, conflict

and power structures, together with their relationship to human resource policies and job design:

- Organisation theorists (e.g., Burns and Stalker, 1961; Woodward, 1965) explore the relationship between environmental and organisational variables. Accounting researchers have expanded the scope of these variables to include relationships between the environment (e.g., technology, uncertainty), the organisation (e.g., structure, task complexity, decentralisation, supervisory style, job-related tension) and accounting variables (e.g., performance evaluation, budgetary participation). The aim is to produce a contingency theory that 'must identify specific aspects of an accounting system which are associated with certain defined circumstances and demonstrate an appropriate matching' (Otley, 1980, p. 414). Such studies have established relationships between accounting practices and environmental and organisational factors, and have produced a recognised body of knowledge (see Otley, 1984), which is arguably the most coherent in management accounting.
- Berry et al. (1985) adopt a case-study approach based in sociological and conflict theories to explore the complexity of organisational processes in the National Coal Board.
- Other organisation theorists (e.g., Ouchi, 1977; Perrow, 1970, 1972; Thompson, 1969) have provided alternative frameworks for systems theory that have facilitated the discussion of the roles of control structures and subordinate behaviour. This has allowed developments in accounting research on budgetary control devoted to the distortion of accounting information systems (e.g., Birnberg et al., 1983).

Alternative research methods are consistent with these different theoretical approaches for individual, group and organisation focus. Thus experimental methods (based on behavioural and psychological expectations) will most frequently be used to measure individual behaviour; survey methods can be used to reveal self-reported attitudes and preferences, and case studies used to explore organisational change. However, a note of caution is warranted in that a consistent factor of the studies in the organisational behaviour area is their focus on relatively few dependent variables: productivity, absenteeism, job turnover and job satisfaction predominate to the exclusion of almost anything else. Staw (1984), among others, has argued for more attention to be devoted to new dependent variables, such as job stress, innovation and individual dissent. Briers and Hirst (1990, p. 374), in their review article, show that the contingency theory literature in accounting has followed a very similar line, in that only four major dependent variables (dysfunctional behaviour, job performance, budgetary performance and unit performance) have been the subject of study. The parallels reveal how dependent we are, as accounting researchers, on the theoretical developments in related disciplines.

Searching the literature

Our search for a research idea will have lead us to the key motivating literatures in the area. Keyword searches in relevant databases will have yielded numerous seemingly relevant abstracts, and a smaller number of pivotal papers. We need to drill down further from these papers, and successive papers, by attending to their reference lists and to the references from the references. There is still no easy way to do this because online databases frequently terminate around 1988, making hardcopy journal articles essential, either through library serial holdings or through inter-library loans. The result of such a search will be a great deal of paper, which should disclose:

- those few key papers, the seminal literature in the area, which will motivate further research;
- evidence of what we know (i.e., the current boundaries of our knowledge) and, just as important, what we do not yet know, because it still constitutes future research (the nature of empirical findings will also show what we are not sure of because of inconsistent or contradictory findings);
- an indication of the theoretical frameworks which are likely to guide future research in the area, and the implications of these for our own research;
- subsequent searches of recently completed theses, conference presentations and working papers, which will reveal both the avenues of research currently being explored (likely two years ahead of publication) and, importantly, those gaps which are still apparent;
- the theories which will indicate the relationships we are able to justify. The empirical literature will have measured associations to identify those variables that are likely to be influential, which are therefore the leading candidates to appear in our early conceptual models.

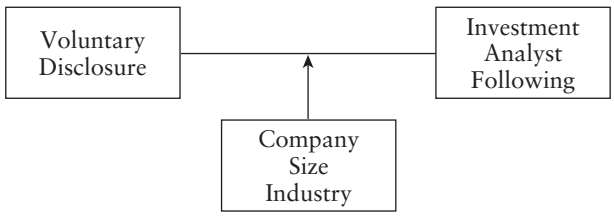
More detailed aspects of the literature review are considered in Chapter 9 (with respect to archival searches) and Chapter 10 (with respect to thesis writing).

We began to consider variable definition and measurement issues in Chapter 2, but it is opportune at this stage to return to Figure 2.6 to consider alternative, perhaps more appropriate, configurations for our conceptual schema.

Modelling the relationship

Our fundamental relationship is an association between two variables of interest, with the relationship subject to influence by another group of variables.

FIGURE 3.2



Voluntary disclosure and reciprocal causality

At this stage we have only an 'association'; we cannot postulate a causal direction without some underpinning theory. We potentially begin with a reciprocal causality – where dependent and independent variables are conceivably reversible in their roles – until rival hypotheses can be eliminated. The scenario described earlier in Chapter 1, with respect to levels of voluntary disclosure, and depicted in Figure 3.2, fits this situation.

There is a relationship between levels of voluntary disclosure by a company and the number of investment analysts who are following (analysing, reporting and issuing recommendations on) that company. A causal link in either direction is feasible, with each supported by a rival explanation:

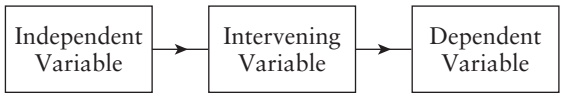
- firms are 'signalling' their practices, innovativeness and early adoption by choosing to disclose additional information voluntarily; or
- firms are subject to the demand for more information from analysts and are responding to the analysts' and the company's mutual benefit.

It is unlikely that both explanations are correct, and we await theory disconfirmation to establish an appropriate causal direction for the relationship. Co-variation is in evidence, but one of the events likely precedes the other. Evidence from Lang and Lundholm (1996), in the USA, suggests that increases in disclosure quality are followed by higher levels of analyst following. However, Walker and Tsalta (2001), with UK data, report contradictory findings of an increase in page length following an increase in analyst forecasting activity, suggesting that the controversy remains.

A further consideration is the nature of the 'influential' variables in a model. A tighter specification of their precise role will be helpful because it will determine both how we manage their control and how we subject the subsequent findings to statistical analysis. Where the influential variables can be thought to be 'intervening', we have the model shown in Figure 3.3.

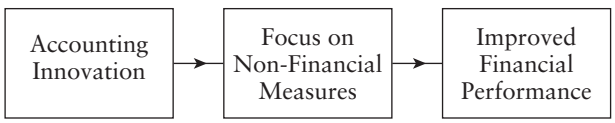
There is no longer a direct relationship between the original independent and dependent variables. Instead, the intervening variable is effectively the dependent variable in one (left-hand) relationship and the independent

FIGURE 3.3



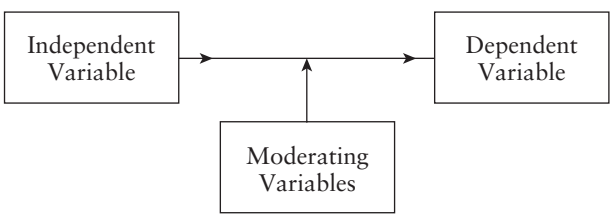
Intervening variables and causality

FIGURE 3.4



Non-financial focus as an intervening variable

FIGURE 3.5



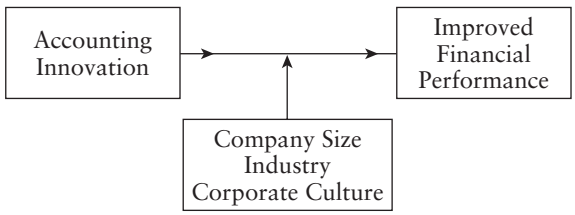
Moderated causal relationship

variable in a second (right-hand) relationship. Figure 3.4 illustrates a possible example of this arrangement in that the introduction of a particular accounting innovation impacts on the adoption and monitoring of non-financial performance measures. It is this new focus, rather than the original innovation itself, that causes improvement in overall financial performance. The implication for statistical analysis of such a model is that we should adopt a partial regression approach.

Alternatively, the influential variables may have a 'moderating' effect in that the relationship between dependent and independent variables is conditional on the values assumed by the other variables, in the manner of Figure 3.5.

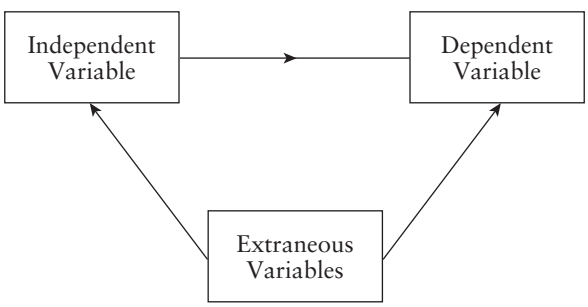
For the situation described above, the impact of accounting innovation may be conditional on the moderating variables of Figure 3.6.

FIGURE 3.6



Size, industry and culture moderating performance

FIGURE 3.7



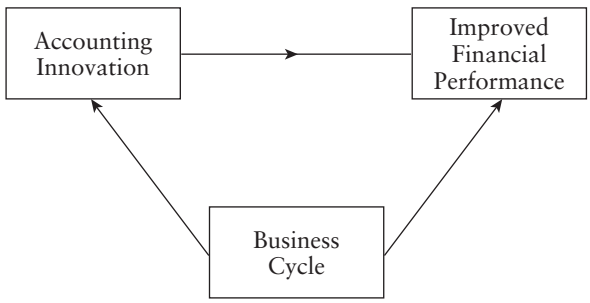
Influence of extraneous variables

Thus the strength of the relationship may depend on the size of the company, for example, so that the association is conditional on companies being larger than a certain level. Where dichotomous variables are being used to reduce measurement error (i.e., large/small companies) then Moderated Regression Analysis (MRA) is the most appropriate form of analysis (see Hartmann and Moers, 1999, for a fascinating description and critique of the technique) and is commonly used within contingency modelling studies (e.g., Dunk, 1993; Mia, 1989).

The influential variables may be superfluous to the relationship of interest because they impact on each separately by exerting an overriding influence. They are then termed 'extraneous' and are shown in Figure 3.7.

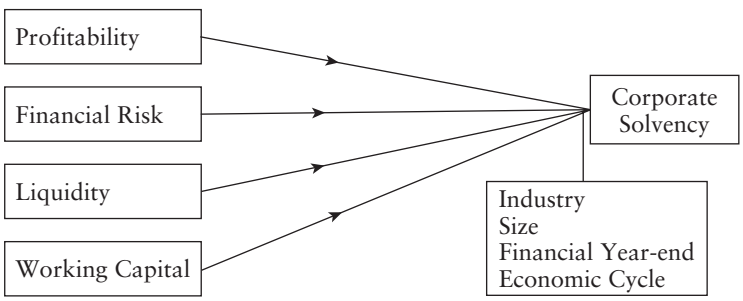
For example, economic conditions may dictate the nature of the relationship. Both levels of accounting innovation and improved levels of financial performance may be linked to the up-swing of the business cycle. To examine any relationship between the variables of interest, either we should examine partial correlations or we should control for the effects of the business cycle by confining our data collection to a single, specified level of economic

FIGURE 3.8



Economic conditions as extraneous variables

FIGURE 3.9



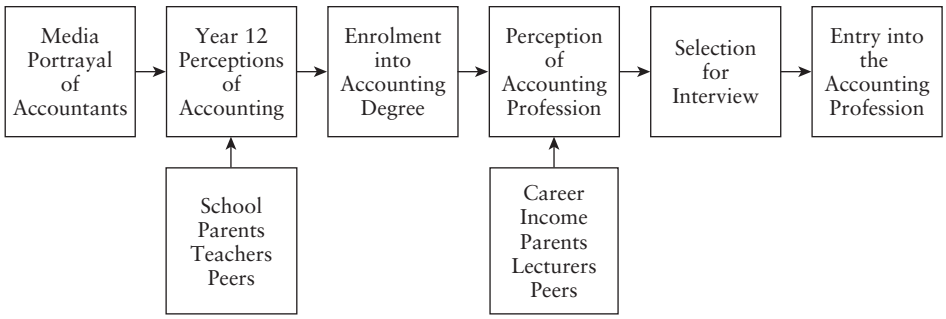
Multiple independent variables

conditions, as suggested in Figure 3.8. Naturally, in doing so our findings may threaten external validity considerations because relationships may only hold for the particular economic conditions for which sample data are sought.

These simple models may be further complicated by the generation of multiple explanatory variables, whose simultaneous effect needs to be examined. For example, the explanation of business failure (e.g., Taffler, 1983), as illustrated in Figure 3.9, would normally be associated with dependence on at least four financial variables (measuring profitability, risk, liquidity and working capital) with the effect moderated by other company-specific characteristics. Even so, a further model would be needed to explain the 'timing' of failure.

If we return belatedly to our example of Chapter 2, we find that we now have a much more complex model than originally initiated, and one which, as Figure 3.10 suggests, may require two separate studies.

FIGURE 3.10



Modelling the recruitment process

In each of the above cases, the precise nature of the analysis that we can perform will be dictated by how we measure the variables of interest. Issues of measurement and testing are discussed in detail in Chapter 4 as part of the data collection and analysis procedures.

Developing the hypotheses

Hypotheses must be testable. Their content must be measurable in some way even if they are not directly observable. For these purposes, 'ratio' (multiplicative) and 'interval' (continuous) scales of measurement are preferred because they make possible a wider number of analytic alternatives, but ordinal (i.e., involving ranks) and nominal (in particular dichotomous, yes/no) scales are common in the accounting literature, and methods exist for their analysis. Once a research question consistent with theory has been formulated, and the research design specified, we need to develop one or more hypotheses for testing. Theory and existing literature should drive the formation of hypotheses so that what we postulate is eminently feasible based on the existing evidence. This is always something of a jump because we are venturing from the known (existing empirical findings) to the unknown (what we are investigating). It is easy to feel uncomfortable about this jump because even where all relevant literature has been digested, the move to hypotheses may still seem large and vulnerable. Where authors are particularly sensitive about this jump they may opt to establish 'propositions' of what may be anticipated rather than formal hypotheses in the form of testable expectations.

The hypotheses will normally be stated in null or alternative forms, and most reviewers prefer the null hypothesis form. The null hypothesis (H_0)

postulates the existence of no relationship between the variables of interest; we then attempt to assemble sufficient evidence to suggest that, statistically, the null hypothesis is not a reasonable assumption. If we have no prior evidence to suggest a direction of causality, then we have no alternative but to adopt a null hypothesis format.

The alternative hypothesis (H_1) postulates the existence of a directed (often causal) relationship, and our assembled evidence must show that findings are inconsistent with no significant relationship (the null position). Examples of hypotheses and associated tests of significance are dealt with in more detail in Chapter 4 while addressing quantitative aspects.

In conducting tests of hypotheses we are faced with the possibility of making two errors:

Type 1 error – the rejection of a true null hypothesis
Type 2 error – the acceptance of a false null hypothesis.

In a legal scenario, the conviction of an innocent man would constitute a Type 1 error, while freeing a guilty man would provide a Type 2 error. In a bankruptcy prediction environment, the misclassification of a failed company as healthy would constitute a Type 1 error, while the classification of a healthy company as a failed one would constitute a Type 2 error. It is generally more important to reduce the probability of Type 1 errors (since they are seen as more serious or more expensive) so that hypothesis testing places more emphasis on Type 1 rather than Type 2 errors. Reducing the level of Type 2 errors would normally involve a trade-off for more Type 1 errors – a trade-off which may be unacceptable. Thus, in the bankruptcy prediction environment, Type 1 errors are virtually unknown, but Type 2 errors are plentiful. It remains a challenge to accounting researchers to reduce the level of Type 2 errors while maintaining current levels of Type 1 error.

If a hypothesis is not supported, we must consider the possibility of competing explanations for the findings. We must also be sceptical of our own findings, question where inconsistencies may have arisen, and be prepared to collect more data or replicate the study.

Validity concerns

These concerns are covered in detail with respect to each of the methods examined in Chapters 6–9, but an early impression of the overall problem is helpful prior to discussion of data collection. The trade-off between reliability and construct validity has been referred to above; that between internal/external validity needs further embellishment. Internal validity threats are confined to what can go wrong during the research. Potential errors in, or changes to, the

measurement instrument may dictate the use of a single version of the instrument to overcome **instrumentation** concerns, but the nature of the manipulation, or ordering concerns, may make this unfeasible. Similarly, **selection** problems in both the attraction of participants and their assignment to groups may be unrepresentative. The passage of time causes particular concerns in longitudinal studies, or in studies where data are collected at more than one time, through drop-outs (**mortality**), fatigue or waning enthusiasm (**maturation**), non-comparability of materials (**history**), and the effects of serial testing where repeated measures are being used (**testing**). These concerns suggest conducting the research over a short time period and at one visit, but this may conflict with the requirements of the research question under study, necessitating more complex remedies. Random selection of subjects is almost unknown in accounting research, so we need to try to introduce randomness at subsequent stages in the process, while performing supporting measurement of potentially influential demographics (e.g., age, gender, experience).

External validity requires that research findings have implications for other sites (companies and countries) and people at different time periods. Unfortunately, the highest levels of internal validity are associated with artificial conditions (the findings may be restricted to the time, cases, participants and location of the research setting). Relaxing assumptions there to introduce greater external validity will inevitably threaten internal validity, and this remains a challenge to accounting researchers.

Where we have solved our variable definition and control methods, we still have to make an informed choice as to the most appropriate research method to be employed. These fall into five broad categories:

- scientific reasoning and/or model building;
- historical research using archival data and/or secondary sources;
- case studies requiring extensive exploration in the field;
- surveys, often involving large-scale sampling, though lacking control and richness of outcome; and
- experiments, either in the field or, more usually, in laboratory-type conditions.

Issues of validity (both internal and external) are central to this choice, and they are considered with respect to the alternative methods in each of the succeeding chapters. Some form of trade-off is usually anticipated between control and structure, on the one hand, and real-world application on the other. Whereas the impact of particular variables (internal validity) will be much clearer under highly structured and tightly controlled conditions, the artificiality of laboratory conditions will reduce the opportunity for generalisation (and external validity). Experimental conditions necessitate the use of situations less complex than those encountered in the real world, with fewer variables and lower information content; most experiments would be impossible to conduct otherwise. The key question remains whether a simplistic situation still includes sufficient elements of reality for it to be a realistic predictor of actual decision-making environments.

4

Data Collection and Analysis

CHAPTER CONTENTS

This chapter is concerned with the collection and analysis of qualitative and quantitative data for use in the subsequent methods described in Chapters 6 (experimental), 7 (surveys), 8 (fieldwork) and 9 (archival). The emphasis is unapologetically on quantitative methods since I see so many research candidates adopting (arguably, more difficult) qualitative methods mainly because they are not confident with numerical methods. One of the primary aims here is to reduce levels of confusion surrounding alternative quantitative methods by demonstrating the circumstances in which particular statistical methods are the most appropriate. Simple numerical examples will be used throughout, though the actual number crunching will almost always be carried out with the help of SPSS and Excel, or similar packages.

A logical sequence is observed in that we first consider sample selection and measurement issues, before proceeding to questions of data analysis, descriptive statistics and, where appropriate, significance testing. As we shall see, many of the problems which can arise at the analysis and testing stages do so because too little attention has been paid earlier to questions of sample selection and measurement.

Sample selection

Although random selection of samples is usually deemed desirable, it may not produce a sample that is either representative or 'useful'. Thus random sampling of companies may not give us any representatives of a particular industry (in failure prediction studies it would quite likely give us few, if any, actually failed cases). A systematic sampling approach is often preferred in practice (e.g., choosing every twentieth item, say, from the sampling frame) but this may still not solve the problems alluded to above. Directed sampling obtained from specified groupings, perhaps random or systematic at this lower level, may therefore be preferred in practice. We are usually violating the stated assumptions of statistical techniques to some extent straight away, though often without causing great harm. As we recognised earlier, we may even be forced to advertise to raise a sample, but in any event most samples that we achieve are voluntary (for experimental conditions) and convenient (for field studies), so they are scarcely random.

The determination of the most appropriate sample size is largely a cost–benefit exercise. We want the biggest sample size that we can afford to collect, in terms of both time and money. Larger sample sizes are usually more expensive in every aspect of performing the research. The representativeness of the research findings and their statistical significance will generally increase as the sample size increases, in all but the most exceptional circumstances. At the very least we should have a sample size large enough to allow us to conduct the required tests of the research question, and we should be aware of this minimum requirement at the outset.

Issues of sample selection are fundamental to our subsequent choice of research method and the most appropriate form of hypothesis testing. Parametric tests of significance necessitate the assumption of independent random samples drawn from normally distributed populations. These conditions are rarely satisfied in practice, so, strictly, we should be choosing the non-parametric alternatives. Similarly, the experimental methods of Chapter 6 require the random selection and assignment of subjects to experimental treatments. Where we are unable to achieve such random selection standards, then we may not have a 'true' experiment.

Measurement issues

Kidder and Judd (1986) differentiate four types of measurement scale – nominal, ordinal, interval and ratio – each with implications for the most appropriate methods for data analysis. For example, techniques like regression analysis

depend on numerically ordered data and are unsuitable for ordinal variables which convey categories rather than levels (e.g., manufacturing, retailing, financial services). Analysis of variance (ANOVA) is preferred in such circumstances. Confusingly, Likert scales are strictly ordinal variables, but they are usually considered as interval variables for analytical purposes because of their relatively large number of categories (i.e., usually five or seven in practice) and the incidence of averaging.

- **Nominal:** Mutually exclusive and collectively exhaustive categories conveying no ordering message (e.g., male/female; manufacturing/retailing/other). Where these are treated as other than (0, 1) variables, care must be exercised in their use as independent variables in a regression analysis, since ordinary least squares (OLS) analysis will treat them as ordered variables. The most common test of statistical significance for use with this data is the chi-square (χ^2) test. Phi and Cramer's V are the most appropriate measures of association.
- **Ordinal:** Mutually exclusive categories which can be ordered (e.g., large/medium/small). Rank-order methods (e.g., Spearman's 'rho') are the most appropriate measures of association for this data, and measures of statistical significance are confined to non-parametric methods (e.g., the Mann–Whitney U-test). Although the use of parametric methods (e.g., t-test, F-test) is theoretically incorrect for ordinal data, some researchers will still adopt these techniques on the grounds that the difference in outcomes is miniscule. Together, nominal and ordinal variables are often called *non-metric* variables.
- **Interval:** Mutually exclusive ordered categories where specific intervals have the same meaning but no ratio relationship exists (i.e., a score of 2.0 is not double the size of a score of 1.0). Thus a temperature of 30 degrees is not twice as hot as 15 degrees; a Z-score of 4.0 is not four times better than a Z-score of 1.0. The product–moment coefficient of correlation (Pearson's 'r') can be used here to measure association, and parametric tests of significance are employed.
- **Ratio:** Continuous data, where specific intervals have the same meaning, and multiples have the same meaning (e.g., age, height, weight, dollars). Again Pearson's 'r' is an appropriate measure of association and parametric significance tests can be used. Interval and ratio measures, together, are often called *metric* variables.

Many data items in the accounting environment (notably dollar-based variables) are ratio scaled. However, there are few situations in the behavioural and organisational arena where the data are so powerful. This causes problems in accounting research where the mix of variables from different measurement levels precludes the use of optimal analytical techniques. The solution is often the scaling-down of the ratio variables so that they are compatible with the other variables of interest, and the adoption of non-parametric methods of analysis (e.g., asset data are often re-scaled to

high/low or high/low/medium to make it more compatible with behavioural data, like task complexity levels, in contingency studies). We must recognise the loss of measurement power and information content, which accompanies such re-scaling decisions, along with the associated testing restrictions. Even though they are ratio-scaled, many financial variables (notably accounting ratios and stock returns) will rarely be normally distributed, which means that they may violate the assumptions for the use of parametric methods anyway! In such circumstances we may need to turn to non-parametric alternatives like the Mann–Whitney U-test (for two sample situations) and the Kruskal–Wallis test (for multiple sample situations). Both of these tests are discussed later in the chapter.

Within our ordinal level measures alternative rating scales are frequently employed in the accounting literature:

- **Linear (sometimes called graphic scales):** e.g., strongly agree/disagree and points in between these extremes, as long as they are not labelled verbally. But respondents may be unwilling to select the extremes on a continuous scale (i.e., a 1 to 7 scale may induce subjects to generate responses confined to the 2 to 6 range).
- **Itemised (sometimes called categorical scales):** as above, but with labels to denote specific ordered categories, i.e.: Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree. The word 'Neutral' is often replaced in practice by the phrase 'Neither Agree nor Disagree'.
- **Comparative:** measures ask for judgements to be made with reference to specific levels of performance with the objective of providing a comparative base.
- **Multiple-item** (of which the Likert scale is much the most commonly adopted): the numerical scores on the Likert scale permit items measuring the same construct to be added. In so doing it facilitates the investigation of the impact of individual items and sub-groups as well as any incidence of multicollinearity.
- **Semantic differential:** following Osgood et al. (1957), who developed a semantic differential to measure individuals' perceptions of the meaning of different terms. A set of seven-point bipolar scales allows respondents to rate concepts between the extremes of good/bad, passive/active, positive/negative, etc. The method has been employed in the accounting environment (most notably by Houghton, 1987, 1988) to measure understanding of accounting terminology.

Measurement issues are fundamental to the availability of optimal analytical techniques. As we will observe in subsequent sections, the absence of ratio-scaled variables may necessitate the use of non-parametric statistical methods. The non-normality of the variables will violate ordinary least-squares regression assumptions, and the incidence of measurement error may preclude the adoption of our preferred methods for model building.

Data management

Ideally, before we commence the data collection, and certainly before the data analysis stage, we need to address issues concerning data management. We will likely have large volumes of data, perhaps quantitative and qualitative, and possibly emanating from alternative sources (e.g., experiments, surveys and documentary materials). We need to have systems in place early to establish the accuracy of the data, and to establish an audit trail should we need to check the sources and/or content of particular data items. My hoarder's mentality leads me to establish an early rule: 'never throw anything away'. This guideline naturally applies to final versions of experimental instruments and completed survey documents, but can usefully be extended to personal notes, 'post-it' reminders and drafts of both pilot instruments and papers for publication. All of these can act as memory aids should we need to return to source materials in order to establish the precise progress of the research process. Most university ethics protocols for research processes will dictate the safe storage of original research materials, including audio-tapes and associated transcripts, for periods up to seven years.

Prior to data collection we need to establish precise variable definitions and easy-to-remember variable labels. Where we have a complex questionnaire, then we need to establish a coding system for the recording of answers. All of these issues need to addressed early, and solutions be carefully documented. Thankfully, computer-based help is at hand to provide assistance, but we must also guard against the perils associated with advanced technology.

Spreadsheet and database software (e.g., Excel and Access, respectively) are now well developed to handle large volumes of numerical data. They can also be used to conduct statistical descriptions without recourse to more sophisticated statistical packages (e.g., SPSS), and also to construct new variables as transformations of existing variables. But opportunities for error will exist, for example:

- How will the software treat missing data values? If we leave a blank, will it be treated as a zero entry? This needs to be checked.
- How do we check for data entry errors? Ideally, we input the data twice, independently and check for variation, though in practice this can be both expensive and time consuming. Random spot-checks and eye-balling of columns vertically will usually reveal errors because they are most often associated with too many or too few key depressions. Alternatively, we can introduce check columns to highlight data irregularities (e.g., in ratio analysis, computing both the current assets ratio and the quick assets ratio to verify the existence of an intuitive relationship between the two). A quick alternative is to make use of simple descriptive statistics: compute the mean and standard deviation of each column with the formula function,

and establish a range of values three standard deviations either side of this mean which we would expect to embrace all of the data values. Monitoring the columns will then establish which data items lie outside of this range so that we can quickly establish whether they are genuine outliers or data errors. If the latter, we should be able to correct the entry; if the former, we may need to make a decision about the inclusion of wild outliers which could potentially distort the research outcomes.

- How do we avoid data corruption? Keep multiple files in different places and never tamper with the master copy! Anyone who has had to reconstruct a spreadsheet file from scratch will echo these principles. This applies particularly to the use of data-sort routines where it is easy to exclude columns on the far-right of the sheet from the sort procedure, with the potentially disastrous consequences of mismatched results.

The management of narrative data poses particular problems. If we are dealing with relatively small numbers of cases, then manual solutions (e.g., the use of different coloured paper for different narrative sources) may be an efficient means of keeping track of individual contributions and, where appropriate, generalising overall. Large numbers of long narrative transcripts almost necessitate the use of dedicated software (e.g., NUDIST) for classifying and categorising the data and facilitating keyword searches.

The basic principles of care in recording and documenting change are well known to accountants, and should make accounting researchers particularly proficient in this aspect of the research process!

Descriptive statistics

What we are trying to do with statistical testing is essentially very simple. We want to compare an **observed** value with what we **expected** to find, and judge whether this difference is big enough to be attributed to chance or not. If the variation between two samples is bigger than the variation within the samples, then we shall observe a difference which is statistically **significant** – it cannot be attributed to a chance occurrence. But if that inter-sample variation is small, then it is feasible that these samples may be drawn from a common population, with observed 'differences' being just random.

We require a standard by which to judge whether an observed difference is significant or not, and this is where statistics come in. There are a number of distinct stages that we adopt in the test procedure:

- state the null hypothesis;
- identify the most appropriate statistical test for the size and nature of sample, and the measurement scale;
- choose the level of significance (almost always 5% or 1%);

- choose an appropriate test statistic;
- compare the 'observed' and 'expected' values of the variable, and compute the test statistic for the difference between the two;
- look up the critical value of the test statistic for the appropriate statistical distribution and number of degrees of freedom (d.f.), where df = the difference between the size of sample and the number of coefficient values (parameters) that the sample has been used to estimate;
- compare the 'test statistic' with the 'critical value' and come to a decision. Normally, if the 'test statistic' is greater than the 'critical value', then we should reject a null hypothesis of no significant difference.

Statistical tests can be classified into two major categories: *parametric statistics* and *non-parametric statistics*. They differ in the assumptions they make about the underlying distribution of the data under analysis. Parametric statistics require that data be drawn from normal distributions, which are smooth, bell-shaped symmetrical curves, defined by mean and standard deviation measures. Non-parametric statistics make no such assumptions regarding the underlying distribution; they describe relationships in terms of frequencies, rankings and directional signs, rather than means and standard deviations. Parametric tests are the more powerful, so we would normally prefer to use them, but if there is any doubt about the quality of the data, or the underlying assumptions, then we would move to the non-parametric alternative. Although this is technically required, the outcomes are usually no different because standard statistical techniques are incredibly robust in practice despite the violation of the underlying assumptions.

We may be conducting an essentially descriptive study, with very few numbers involved, but we still have at our disposal a powerful statistical armoury to add to the integrity of our findings. Descriptive studies often record simple proportions, cross-tabulations and measures of association, even where there is no formal hypothesis testing or model building. We want to know if the **observed** values differ significantly from what we would **expect** if, in fact, no relationship existed at all, and simple statistics allow us to draw such inferences. Table 4.1 provides a summary of the statistical tests at our disposal for typical descriptive situations. Examples of each are provided below.

Table 4.1 is inspired by Cooper and Emory (1995, p. 445) but includes only those non-parametric tests in common use in the accounting literature. Cooper and Emory also discuss, among others, the use of the Kologorov–Smirnov test and Runs test (for single-sample ordinal combinations), the Wilcoxon matched-pairs test (for related samples/ordinal) and the McNemar test (for related samples/nominal), none of which is considered further here.

Observed proportions

Consider first the use of a single sample to generate the proportion of observations which meet our requirements. We can use the Z-test of the Normal

TABLE 4.1

Summary of significance tests by measurement level

Measurement level	Single sample	Two samples		Multiple samples
		Independent	Related	Independent
Nominal	χ^2 test	χ^2 test	–	χ^2 test
Ordinal	–	Mann–Whitney U-test	Sign Test	Kruskal–Wallis Test
Interval and ratio	t-test Z-test	t-test Z-test	Paired-case t-test	Analysis of variance (ANOVA)

FIGURE 4.1

OBSERVED PROPORTION [O] 64% based on sample size of n = 100 observations

EXPECTED PROPORTION [E] 50%

for a null hypothesis of no actual difference

i.e., $p = 0.50$
$q = (1 - p) = 0.50$

Mean = $np = 100(0.50) = \underline{\underline{50}}$

Standard Deviation = $\sqrt{npq} = \sqrt{100(0.50)(0.50)}$

$= \underline{\underline{5.0}}$

95% confidence interval for the sample mean:

$50 \pm (1.96)5.0 = 50 \pm 9.80$

i.e., a range from 40.20 to 59.80 is consistent with the null hypothesis

But the Observed Proportion of 64% is outside this range and so is unacceptably large to be consistent with the null hypothesis, i.e., reject the null hypothesis – 64% is a significant finding.

The equivalent Z-test for these data is:

Test statistic: $\dfrac{[0] - np}{\sqrt{npq}} = \dfrac{64 - 50}{5.0} = \underline{\underline{2.80}}$

Critical value: $Z_{0.05} = \underline{\underline{1.96}}$

Since test statistic (2.80) exceeds the critical value (1.96) reject the null hypothesis at a 5% level of significance.

Significance of test of proportion

Distribution to test the significance of the observed proportion detailed in Figure 4.1. Effectively we are asking the question: is a proportion of 64% from a sample of 100 observations big enough to convince us that it is not a chance sample from a population with an actual proportion of 50%?

Cross-tabulations

The use of cross-tabulations to generate contingency matrices represents the most popular of analytical tools. A focus on the individual cells of the matrix

FIGURE 4.2

[O] OBSERVED FREQUENCIES (from Figure 2a in Smith et al., 2001)

		ACCOUNTING POLICY CHANGE		
		NO CHANGES	SOME CHANGES	TOTAL
COMPANIES AUDITED BY	BIG 8	97	118	215
	NON-BIG 8	94	58	152
	TOTAL	191	176	367

a = 2 rows

b = 2 columns

[E] EXPECTED FREQUENCIES (assuming no relationship between auditor/changes)

$\frac{(191)(215)}{367}$	$\frac{(176)(215)}{367}$	111.9	103.1
$\frac{(191)(152)}{367}$	$\frac{(176)(152)}{367}$	79.1	72.9

≡

Test statistic:

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(14.9)^2}{111.9} + \frac{(14.9)^2}{79.1} + \frac{(14.9)^2}{103.1} + \frac{(14.9)^2}{72.9}$$

$$= \underline{9.988}$$

Critical χ^2 value at 5% level of significance = $\chi^2_{(a-1)(b-1)^{0.05}}$

$$= \chi^2_{1,\,0.05} = \underline{\underline{3.84}}$$

χ^2-test for difference in frequencies

allows us to compare the **observed** frequencies with those we would **expect** to see, should there be no relationship between the categories (the null hypothesis). We use a chi-square test (χ^2) usually with a 5% level of significance. In using the χ^2-test no cell should have an expected value which is less than one, and single-figure expected frequencies should be rare (i.e., less than one in five cells). If this is not the case, we will either require a small sample correction or, more likely, need to aggregate some cells by reducing the size of the matrix in order to be able to conduct the test. Figure 4.2 details the conduct of this test, comparing 'observed' and 'expected' numbers of accounting policy changes.

Since the test statistic (9.988) is greater than the critical value (3.84) then we can reject a null hypothesis of no link between auditor grouping and the incidence of accounting policy changes.

Correlation coefficients

Measures of association are dealt with in more detail in the next section, but it is opportune to ask and answer another simple question here: is the observed correlation coefficient big enough for us to be confident that we have a real, not illusory, relationship here? That is, is a coefficient of 0.554 based on a sample of just 17 paired observations big enough for us to recognise a significant relationship? Although it may not appear to be a high measure of association, and the sample size is relatively small, then, as we see in Figure 4.3, it is too big for us to expect by chance.

Differences in sample means

Tests of differences of means between pairs of variables provide a major focus for accounting research. A variety of tests are available depending on the measurement level involved and the way in which the sample has been drawn.

Independent samples

Figure 4.4 illustrates the calculations for two samples (X) and (Y), each of size 17, of ratio-scale data drawn as independent samples from normal populations. The question we want to answer is: are these samples similar enough to each other for us to judge that they could conceivably have been drawn from the same population? We answer this by testing the difference between the two sample means, relative to the variation that exists within the samples. If the difference between the samples is greater than that within

FIGURE 4.3

The data of Figure 4.7 (see page 70) shows a correlation coefficient of r = 0.554 calculated from a sample size of n = 17

Mean (r) = 0.554

$$\text{Standard deviation} = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - (0.554)^2}{15}} = 0.2150$$

$$\text{Test statistic} = \frac{\text{Mean}}{\text{Standard deviation}} = \frac{0.554}{0.2150} = 2.577$$

$$\text{Critical t-statistic} = t_{n-2,\,0.05} = t_{15,\,.05} = 2.131$$

Test statistic is greater than the critical value so we should reject a null hypothesis of zero population correlation coefficient.

[Equivalent 95% confidence interval: 0.554 ± (2.131)(0.2150) giving a range of 0.096 to 1.012 which does not include r = 0 as a possible value.]

Test of significance of correlation coefficient

the samples, then we would infer that the samples are drawn from different populations, and reject the null hypothesis.

Paired cases

Where the samples drawn are not independent but represent a before-and-after situation involving the same subjects (usually people), then we have a repeated measures situation for which more powerful statistical tests are available. The paired-case t-test effectively controls for individual differences between subjects so that the observed differences are wholly attributable to the treatments (changed conditions) assuming that there are no carry-over effects between the two conditions. In this case we can observe non-zero differences between the two samples, but are these differences in the same direction and big enough not to be attributed to chance? Figure 4.5 details the measurement (in column 3) of the differences in scores between the two samples, allowing the calculation of the mean and standard deviation for the d-score. A null hypothesis suggests a population mean of zero and a test statistic of:

$$\text{Test statistic:} \quad \frac{d - 0}{\sqrt{\frac{s_d^2}{n - 1}}} = \frac{(4)(4)}{\sqrt{62.5}} = 2.02$$

Critical value: $t_{n-1,0.05} = t_{16,0.05} = 2.12$

FIGURE 4.4

(1) X	(2) Y	(3) X^2	(4) Y^2
13	16	169	256
14	18	196	324
15	19	225	361
17	21	289	441
18	20	324	400
26	27	676	729
28	17	784	289
29	24	841	576
30	12	900	144
27	23	729	529
31	11	961	121
32	22	1024	484
33	23	1089	529
34	35	1156	1225
36	25	1296	625
37	38	1369	1444
39	40	1521	1600
$\sum X = 459$	$\sum Y = 391$	$\sum X^2 = 13549$	$\sum Y^2 = 10077$

$n_1 = n_2 = 17$

$N = (n_1 + n_2) = 34$

MEANS

$$\bar{Y} = \frac{\sum Y}{n_2} = \frac{391}{17} = \underline{23}$$

$$(\overline{X + Y}) = \frac{\sum X + \sum Y}{n_1 + n_2} = \frac{850}{34} = 25$$

SAMPLE VARIANCES

$$S_X^2 = \frac{\sum X^2 - n_1 \bar{X}^2}{n_1 - 1} = \frac{13549 - 17(27)^2}{16}$$

$$= \underline{72.25}$$

$$S_Y^2 = \frac{\sum Y^2 - n_1 \bar{Y}^2}{n_2 - 1} = \frac{10077 - 17(23)^2}{16}$$

$$= \underline{67.75}$$

$$\text{Test statistic} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n_1 - 1} + \frac{S_Y^2}{n_2 - 1}}} = \frac{27 - 23}{\sqrt{\frac{72.25}{16} + \frac{67.75}{16}}} = \underline{1.352}$$

Critical value $t_{n_1 + n_2 - 2,\ 0.05} = t_{32,\ 0.05} = \underline{2.04}$

Since test statistic (1.352) < critical value (2.04) we cannot reject a null hypothesis of insignificant difference between the groups at a 5% significance level.

t-test for difference in means

Despite the increased power of the test in the paired-case situation, the test statistic still does not exceed the critical value, so the null hypothesis cannot be rejected at the 5% level of significance.

Non-parametric alternatives

Where the sampling is not random, the resulting samples will be potentially unrepresentative. If assumptions about the population distribution are unwarranted, then parametric tests may be unreliable and we should adopt non-parametric tests. If we return to the data of Figure 4.5, but regard the

FIGURE 4.5

(1) X	(2) Y	(3) d = (X – Y)	(4) d^2
13	16	–3	9
14	18	–4	16
15	19	–4	16
17	21	–4	16
18	20	–2	4
26	27	–1	1
28	17	11	121
29	24	5	25
30	12	18	324
27	23	4	16
31	11	20	400
32	22	10	100
33	23	10	100
34	35	–1	1
36	25	11	121
37	38	–1	1
39	40	–1	1
		$\sum d = 68$	$\sum d^2 = 1272$

$n = 17$

MEAN

$$\bar{d} = \frac{\sum d}{n} = \frac{68}{17} = \underline{4}$$

SAMPLE VARIANCE

$$S_d^2 = \frac{\sum d^2 - n\bar{d}^2}{n_1 - 1} = \frac{1272 - 17(4)^2}{16} = \underline{62.5}$$

t-test for paired-case difference in means

data as ordinal, perhaps because of potential measurement error which may distort the outcome of conventional t-tests, then we can use the **Sign Test** as a non-parametric alternative (for the before-and-after paired situation) and the **Mann–Whitney U-test** (where we have independent samples).

The Sign Test

We simply monitor column 3 of Figure 4.5 to determine the number of positive and negative differences we have in the paired cases. For the 17 pairs, we observe nine negative and eight positive differences; there are no ties (i.e., no instances of no-change cases), which would otherwise be eliminated from the analysis.

A null hypothesis would lead us to expect 17/2 = 8.5 signs of each direction Our test statistic would be:

$$\text{Test statistic: } \frac{\text{Observed} - \text{Expected}}{\text{Standard deviation}} = \frac{9-8.5}{\sqrt{17\left(\frac{8.5}{17}\right)\left(\frac{8.5}{17}\right)}} = \underline{0.243}$$

Critical value = $Z_{0.05}$ = 1.96

FIGURE 4.6

(1) X	(2) Y	(3) R_X	(4) R_Y
13	16	32	29
14	18	31	25.5
15	19	30	24
17	21	27.5	22
18	20	25.5	23
26	27	16	14.5
28	17	13	27.5
29	24	12	18
30	12	11	33
27	23	14.5	19.5
31	11	10	34
32	22	9	21
33	23	8	19.5
34	35	7	6
36	25	5	17
37	38	4	3
39	40	2	1
		$\sum R_X = 257.5$	$\sum R_Y = 337.5$

$n = 17$

$m = 17$

MEAN

$$\bar{U} = \frac{n}{2}(n + m + 1)$$

$$= \frac{17}{2}(35) = \underline{297.5}$$

VARIANCE

$$S_u^2 = \frac{nm(n + m + 1)}{12}$$

$$= \frac{(17)(17)(35)}{12}$$

$$= \underline{842.92}$$

Mann–Whitney U-test for difference in means

Clearly the test statistic (0.243) is well below the critical value (1.96) so again we cannot reject the null hypothesis.

The Mann–Whitney U-test

Figure 4.6 details the conduct of the Mann–Whitney U-test, which is based on the rank order of the sample values.

The test statistic is:

$$\text{Test statistic: } \frac{R_x - \bar{U}}{\sqrt{S_U^2}} = \frac{257.5 - 297.5}{\sqrt{842.92}} = \underline{1.38}$$

Critical value = $Z_{0.05} = \underline{1.96}$

So that once again the null hypothesis cannot be rejected; the critical value (1.96) comfortably exceeds the test statistic (1.38).

Strictly speaking, to employ this version of the Mann–Whitney U-test one of the two samples (n or m) should be greater than 20 to invoke the Normal approximation. The proximity of our sample sizes (n = m = 17) suggests

that it will make little difference, but for technical accuracy we will compute the small sample test:

$$U_X = nm + \frac{n(n+1)}{2} - R_X = (17)(17) + \frac{(17)(18)}{2} - 257.5 = 184.5$$

$$U_Y = nm + \frac{n(n+1)}{2} - R_Y = (17)(17) + \frac{(17)(18)}{2} - 337.5 = 104.5$$

We compare the smallest of these two U values ($U_Y = 104.5$) with a critical value taken from Mann–Whitney tables (87 for $n = m = 17$). In this case the test statistic must be *smaller* than the critical value in order to reject the null hypothesis, so since 104.5 exceeds 87 we still cannot reject the null hypothesis.

Measures of association

The most appropriate measure of association is again determined by the measurement level, as detailed in Table 4.2. Each of the measures in the table are considered below.

Pearson correlation coefficient (r)

The correlation coefficient matrix is, arguably, the single most useful piece of preliminary diagnostic information. It serves three vital functions:

- it establishes the direction of any relationship which should be intuitively correct and which must correspond with the sign of this variable in any regression equation;
- it suggests those variables which are likely to be useful explanatory variables because they are highly correlated with the dependent variable;
- it highlights potential multicollinearity problems by quantifying the strength of association between competing explanatory variables.

Figure 4.7 illustrates the calculation of the coefficient for interval/ratio data so that it yields a value in the range +1 (for perfect positive correlation) and –1 (for a perfect inverse relationship).

The statistical significance of this coefficient was tested earlier (in Figure 4.3). The square of this correlation coefficient is called the Coefficient of Determination (R^2), and indicates the percentage of the variation in one variable explained by changes in the other. Thus for $r = 0.554$, $R^2 = 0.307$, i.e., 30.7% of the variation in Y is explained by changes in X (and vice versa). The other 69.3% of variation is currently unexplained, and will remain so until we

TABLE 4.2

Measures of association by measurement level

Measurement level	Measure of association
Nominal	Phi Cramer's V
Ordinal	Spearman's 'rho' (Coefficient of Rank Correlation)
Interval and ratio	Pearson's 'r' (Product–moment Coefficient of Correlation)

FIGURE 4.7

(1) X	(2) Y	(3) XY	(4) X^2	(5) Y^2
13	16	208	169	256
14	18	252	196	324
15	19	285	225	361
17	21	357	289	441
18	20	360	324	400
26	27	702	676	729
28	17	476	784	289
29	24	696	841	576
30	12	360	900	144
27	23	621	729	529
31	11	341	961	121
32	22	704	1024	484
33	23	759	1089	529
34	35	1190	1156	1225
36	25	900	1296	625
37	38	1406	1369	1444
39	40	1560	1521	1600
$\sum X = 459$	$\sum Y = 391$	$\sum XY = 11177$	$\sum X^2 = 13549$	$\sum Y^2 = 10077$

$n = 17$

MEANS

$$\bar{X} = \frac{\sum X}{n} = \frac{459}{17} = \underline{27}$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{391}{17} = \underline{23}$$

SAMPLE VARIANCES

$$S_X^2 = \frac{\sum X^2 - n\bar{X}^2}{n_1 - 1} = \frac{13549 - 17(27)^2}{16} = \underline{72.25}$$

$$S_Y^2 = \frac{\sum Y^2 - n\bar{Y}^2}{n - 1} = \frac{10077 - 17(23)^2}{16} = \underline{67.75}$$

$$\text{Pearson's } r = \frac{\sum XY - n\overline{XY}}{\sqrt{(n-1)^2 \cdot S_X^2 \cdot S_Y^2}} = \frac{(11177) - 17(27)(23)}{\sqrt{(256)(72.25)(67.75)}} = \underline{0.554}$$

Product–moment correlation coefficient (Pearson's r)

FIGURE 4.8

(1) X	(2) Y	(3) R_X	(4) R_Y	(5) d $(R_X - R_Y)$	(6) d^2	n = 17
13	16	17	15	2	4	
14	18	16	13	3	9	
15	19	15	12	3	9	
17	21	14	10	4	16	
18	20	13	11	2	4	
26	27	12	4	8	64	
28	17	10	14	–4	16	
29	24	9	6	3	9	
30	12	8	16	–8	64	
27	23	11	7.5	3.5	12.25	
31	11	7	17	–10	100	
32	22	6	9	–3	9	
33	23	5	7.5	–2.5	6.25	
34	35	4	3	1	1	
36	25	3	5	–2	4	
37	38	2	2	0	0	
39	40	1	1	0	0	
					$\sum d^2 = \underline{\underline{327.5}}$	

For n = 17

$$\rho \text{ (rho)} = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6(327.5)}{17(288)} = \underline{\underline{0.5987}}$$

Coefficient of rank correlation (Spearman's rho)

have the opportunity to build multivariate causal models surrounding these variables.

Spearman coefficient ('rho')

As before, where we are unsure of the quality of the data, or of the populations from which they are drawn, we prefer to use rank-order methods (see Figure 4.8).

The conversion of continuous data to ranks (and the corresponding move from Pearson's r to Spearman's rho) is convenient if we are suspicious of the presence of measurement error. Rho is unaffected by data transformations and deals equitably with problems associated with outliers, without the necessity for winsorizing the data, that is eliminating extreme observations from the data which would otherwise bias the outcomes of the analysis. Its

FIGURE 4.9

For the 2 * 2 contingency table of Figure 4.2

where

a	b
c	d

≡

97	118
94	58

and $\chi^2 = \underline{9.988}$ and $N = (a + b + c + d) = \underline{367}$

$$\underline{\text{Cramer's V}} = \sqrt{\frac{\chi^2}{N^* \min\,[r-1,\, c-1]}} \quad \text{for r, rows – c, columns}$$

$$= \sqrt{\frac{9.988}{367 * 1}} = \underline{\underline{0.165}}$$

a very low measure of association reflecting large cell differences

$$\underline{\text{Phi Coefficient}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{(97)(58) - (118)(94)}{\sqrt{(215)(152)(191)(176)}}$$

$$= -\underline{\underline{0.165}}$$

a very low inverse measure of association reflecting large cell differences

[Cramer's V and Phi equivalent for 2 * 2 contingency table]

Measure of association within contingency tables

only major deficiency is with regard to tied ranks; too many of these in small samples may distort the size of the coefficient. As is evident from Figure 4.8, the Spearman coefficient is simple to calculate and yields similar answers to the Pearson coefficient.

Phi and Cramer's V

In Figure 4.2 we were able to reject a null hypothesis for the dataset relating auditor grouping to accounting policy changes. The very low measures of association in Figure 4.9 reflect the earlier findings. Where the contingency table reflects an ordered set of categories on each axis, then the chi-square test is marginally less powerful than alternatives (e.g., Goodman and Kruskal's *gamma* statistic or Kendall's *tau* statistic). Neither of these is considered further here, except to note that Table 2c of Smith et al. (2001) makes use of the latter (see Appendix 2).

Analysis of variance

If we return to Table 4.1 briefly, we note that there are two tests that we still need to address – the analysis of variance (ANOVA) and the Kruskal–Wallis test – for multiple independent samples of, respectively, ratio and ordinal data. We consider a three-sample illustration in Figure 4.10.

Three samples, each of size $n_1 = n_2 = n_3 = 10$ are drawn (initially) from normal populations. The question we want to answer is: could these three samples conceivably have been drawn from the same population, or are the differences between them too large for that to be realistic? We could address the question by conducting separate t-tests between each pair of samples, but that would not be an efficient method, nor an appropriate one were the samples not to prove independent. We therefore conduct a single test – a one-way analysis of variance – to determine whether the variation between the three samples is greater than the variation evident within the samples.

The analysis of Figure 4.10 demonstrates that the between-sample variance is sufficiently greater than the within-sample variance for us to reject the null hypothesis and infer that the samples are indeed drawn from different populations.

ANOVA (analysis of variance) was introduced to the accounting literature by Ashton (1974) as a model for measuring the significance, and percentage variance, explained by the main effects of treatments and interactions between treatments. It was immediately widely adopted as a means of eliminating multicollinearity problems, heteroscedasticity problems when group sizes are equal, and facilitating an unbiased estimation of both main and interactive effects. In theory, the size of the factorial design can be expanded greatly, but in practice it quickly becomes unmanageable; frequently third and fourth order interactive effects cannot be explained satisfactorily, limiting their contribution to the development of theory. The dichotomisation of data necessary to conduct ANOVA reduces the effects of measurement errors, but it may incur information loss.

Table 4.3 details the manner in which ANOVA results would typically be presented. It includes the sources of variation, the degrees of freedom, sums of squares, mean squares and the calculated value of the F-test statistic for the data of Figure 4.10.

If we now relax our assumptions, we move to the case detailed in Figure 4.11. We are still operating with the three-case sample of Figure 4.10, but this time our doubts about the quality of the data dictate that we treat the data as ordinal, rather than ratio, and employ non-parametric methods. This means we need to use the rank order of the data, rather than actual values, and the Kruskal–Wallis test. First all of the sample observations are ordered (from 1 to 30), taking care to average appropriately where ties occur. We then sum the ranks appropriate to each sample and compute the test statistic (see Figure 4.11).

FIGURE 4.10

TREATMENT OUTCOMES

SAMPLE (Y_1) ($n_1 = 10$)	SAMPLE (Y_2) ($n_1 = 10$)	SAMPLE (Y_3) ($n_1 = 10$)	TOTAL OBSERVATIONS $n_1 + n_2 + n_3 = N = 30$
32	35	44	
30	38	46	
35	37	47	
33	40	47	
35	41	46	
34	35	43	
29	37	47	
32	41	45	
36	36	48	
34	40	47	
$\sum Y_1 = 330$	$\sum Y_2 = 380$	$\sum Y_3 = 460$	$\sum(Y_1 + Y_2 + Y_3) = 1170$
Means: $\bar{Y}_1 = 33$	$\bar{Y}_2 = 38$	$\bar{Y}_3 = 46$	$\bar{Y} = 39$
Variance: $S_1^2 = 4.6$	$S_2^2 = 5.0$	$S_3^2 = 2.2$	

where $S^2 = \dfrac{\sum(Y_i - Y_i)^2}{n_i}$ for each of i = 1, 2, 3

Between-sample variance $= \dfrac{\sum n_i(\bar{Y}_i - \bar{Y})^2}{K - 1}$ for K = no. of samples

$$= \frac{10(33 - 39)^2 + 10(38 - 39)^2 + 10(46 - 39)^2}{3 - 1} = \underline{\underline{430}}$$

$$\text{Within-sample variance} = \frac{\sum n_i S_i^2}{N - K} = \frac{10(4.6) + 10(5.0) + 10(2.2)}{30 - 3} = \underline{\underline{4.37}}$$

$$\text{Test statistic} = \frac{\text{Between-sample variance}}{\text{Within-sample variance}} = \frac{430}{4.37} = \underline{\underline{98.4}}$$

Critical F-value $= F_{K-1,\, N-K,\, 0.05} = F_{2,\, 27,\, 0.05} = \underline{\underline{3.37}}$

(i.e., 5% level of significance, with 2 (nominator), 27 (denominator) degrees of freedom)

Since the test statistic (98.4) is greater than the critical value (3.37) we must reject the null hypothesis. There is a significant difference in the outcomes from the three treatments.

One-way analysis of variance (ANOVA)

TABLE 4.3

ANOVA summary table

Source of variation	Degrees of freedom	Sums of squares	Mean square	Test statistic
Model	2	860	4.30	98.4
Residual error	27	118	4.37	
Total	29			

Once again our test statistic exceeds the critical value, leading us to reject the null hypothesis and infer that the samples are indeed drawn from separate populations.

Multivariate model building

Depending upon the measurement level of the data and the role played by the associated variables (whether dependent or explanatory), a variety of model-building methods will be available, as detailed in Table 4.4.

Table 4.4 is adapted from the flow chart provided by Cooper and Emory (1995, p. 521). Conjoint Analysis and Canonical Analysis are rarely used in the accounting literature (they are mainly to be found in the marketing literature) and so are not discussed further. LISREL (Linear Structural Equations Model for Latent Variables) is omitted from Table 4.4 because its allocation to any cell would be arbitrary. LISREL embraces path analysis and structural equation modelling, among others, and can handle dependent and independent variables which may be nominal, ordinal, interval or ratio-scale measures. It is becoming increasingly popular in the accounting literature, so I shall discuss it briefly below.

Regression analysis

Whereas time series analysis can provide us with trend projections for a key variable, in practice this may not be enough. If we wish to influence future values, through appropriate management action, we need to know which variables impact on the values assumed by the key variable. In essence, we wish to establish the degree of association between variables and any causal

FIGURE 4.11

Y_1	RANK	Y_2	RANK	Y_3	RANK
32	3.5	35	9.5	44	22
30	2	38	16	46	24.5
35	9.5	37	14.5	47	27.5
33	5	40	17.5	47	27.5
35	9.5	41	19.5	46	24.5
34	6.5	35	9.5	43	21
29	1	37	14.5	47	27.5
32	3.5	41	19.5	45	23
36	12.5	36	12.5	48	30
34	6.5	40	17.5	47	27.5

$\sum R_1 = 59.5$ $\sum R_2 = 150.5$ $\sum R_3 = 255.0$

$$\text{Test Statistic} = H = \frac{12}{N(N-1)} \sum \frac{T^2}{n} - 3(N+1)$$

where T = sum of ranks in any column
n = no. of cases in that sample; N = total number of cases (30)
K = no. of samples (3)

$$H = \frac{12}{30(29)} \sum \frac{(59.5)^2 + (150.5)^2 + (255)^2}{10} - 3(31)$$

$$= \frac{12}{870} * \frac{91215.5}{10} - 93 = 125.81 - 93 = \underline{\underline{32.81}}$$

However, where there are a large number of tied ranks (as in this case)

i.e., 7 in sample 1, 9 in sample 2, 6 in sample 3

an adjusting factor 'C' is applied where $C = 1 - \sum \frac{t^3 - t}{N^3 - N}$

$$\text{i.e., } C = 1 - \sum \frac{(7^3 - 7) + (9^3 - 9) + (6^3 - 6)}{30^3 - 30} = 1 - \frac{1266}{26970} = \underline{\underline{0.96}}$$

$$\text{Revised test statistic} = \frac{H}{C} = \frac{32.81}{0.96} = \underline{\underline{34.17}}$$

Critical value $\chi^2_{K-1,\,0.05} = \chi^2_2 = \underline{\underline{5.99}}$

Since the test statistic (34.17) easily exceeds the critical value (5.99) we must again reject the null hypothesis.

The Kruskal–Wallis multiple sample test

TABLE 4.4

Summary of multivariate model-building methods by measurement level

		Independent variables	
		Nominal and ordinal	Interval and ratio
Dependent variables	Nominal and ordinal	Conjoint Analysis Canonical Analysis	Discriminant Analysis Logit and Probit
	Interval and ratio	Multivariate Analysis of Variance (MANOVA) Multiple Regression	Canonical Analysis Multiple Regression

relationships between those variables in order to develop an explanatory relationship which allows us to show how and why key variables are changing.

Alternative forms of regression analysis may be adopted depending on the nature of the causal model:

- ordinary least squares regression (for standard causal relationships);
- ordinary least squares regression with dummy variables (for simple conditional relationships);
- moderated regression (for moderating variables);
- path analysis (partial regression for intervening variables).

For the simple two variable (X,Y) situation, a scatter diagram with Y on the vertical axis and X on the horizontal would reveal the strength of any linear relationship between the two variables. We may speculate on the existence of a linear relationship of the form: $Y = a + bX$. To specify the values of the parameters 'a' and 'b' we need to fit a straight line to the points – effectively averaging out their position and establishing the average to which they regress.

The ordinary least squares (OLS) solution to this problem measures the vertical deviation of points away from a fitted line, and ensures that the optimum fit is such that the sum of the squares of these distances, over all the points, is as small as possible. The fitted line is designated $\hat{Y} = \hat{a} + \hat{b}X$, with 'hats' (^) added to signify that we are dealing with estimates based on a

sample of observations. The line is fitted with reference to the vertical distances ($Y - \hat{Y}_i$) of the points from the line, where $e_i = Y - \hat{Y}_i$ signifies the 'error' involved in fitting.

The OLS regression line is fitted to satisfy simultaneously two conditions:

i. $\Sigma e_i = 0$
   Positive and negative deviations must exactly balance, and
ii. Σe_i^2 is a minimum
   The sum of the squares of the vertical deviations from the line is as small as possible.

The specification of the 'a' and 'b' parameters to minimise Σe^2 can be derived using differential calculus such that:

$$\hat{b} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

for n pairs of observations, and

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

For example, the data of Figure 4.7 would generate a regression equation of the form:

$$\hat{Y} = 37.48 + (0.5363)\ \hat{X}$$

on the assumption that we had a theory to justify changes in Y being caused by changes in X. Fortunately, most spreadsheet software will calculate regression and correlation coefficients, as well as providing a graphic plot of the extent of the linearity, without the user needing recourse to the above formulae!

We know from Figure 4.7 that the Pearson correlation coefficient is: $r = 0.554$, and from Figure 4.3 that this coefficient is significant at the 5% level of significance. However, if we had $r = 0$, it would not necessarily mean that no relationship exists, only that no meaningful *linear* relationship exists. For example, a circular relationship between X and Y would generate a linear correlation coefficient, $r = 0$ even though a perfect non-linear relationship would be in existence.

Similarly, a non-zero correlation coefficient does not necessarily mean that a real linear relationship exists. Two totally unrelated variables will inevitably yield small, but non-zero, spurious correlation coefficients by chance. Statistical tests of significance like those in Figure 4.3 will demonstrate

whether or not such values are small enough to constitute non-zero sample estimates from a zero population.

When we come to perform statistical tests of the significance of the parameter estimates, a rule of thumb for sample sizes in excess of 30 is the ratio:

$$\frac{\text{Sample Estimate}}{\text{Standard error of the estimate (se)}} > 2$$

then the sample estimate is statistically significant (i.e., it is too big to be a chance estimate of a non-existent relationship). Our standard normal ordinate has been 1.96 (just as the standard chi-square ordinate χ^2_1, is 3.84, where this is the square of 1.96). For smaller samples the t approximation is about 2.0, and the standard F-value about 4.0 (since $F = t^2$ for one degree of freedom).

These critical values (2 for the t-test and 4 for the F-test) will each vary depending on the size of sample (n) used, and the number of parameters (k) which the data have been used to estimate. As (n – k) gets smaller, the critical values of the t-test and F-test will increase.

OLS regression methods attempt to estimate the actual relationship $Y_i = a + bX_i + \mu_i$ with an estimated relationship based on a finite sample size of n observations. The error term μ_i in the relationship is estimated by the residual of the equation e_i.

OLS fits make a number of assumptions, the violation of which can result in unreliable equations:

1. μ is a random variable;
2. the mean value of μ is zero;
3. the variance of μ is constant;
4. the variable μ is normally distributed;
5. the random terms from different observations (μ_t, μ_{t-1}) are independent;
6. μ_i is independent of the explanatory variables;
7. the explanatory variables are measured without error;
8. the explanatory variables are not perfectly linearly correlated;
9. any variable aggregation has been carried out appropriately;
10. the identified relationship has a unique mathematical relationship;
11. the relationship has been correctly specified.

For analysis, the assumptions fall conveniently into two groups. The first six assumptions (numbers 1 to 6) concern the error term μ_i, as estimated by the residual term e_i. The last six assumptions (numbers 6 to 11) concern the behaviour of the explanatory variables. Several of the assumptions may be difficult to test, especially when only limited data are available. In practice, the verification of assumptions 2, 3, 5, 6 and 8 is the most critical.

Assumption 2 suggests that $\bar{\mu} = 0$ and assumption 6 that $r_{\mu}.x_i = 0$. If either is violated, then parameter estimates will be both biased and inconsistent,

that is, OLS will generate wrong answers, which will not be improved upon by seeking a larger sample size. In practice, we will fit the regression to ensure that $\bar{e}_i = 0$ so that we need only observe $r_{e^2.x_i}$ to verify assumption 6.

Assumption 3 suggests that s_μ^2 is constant, verified in practice by observing any variance in the estimate $\Sigma e^2/(n - k)$ around a fitted regression equation embracing the estimation of k parameters. If this assumption is not satisfied, then heteroscedasticity exists and formulae for parameter estimates, and hence associated significance tests, may be inefficient. Assumption 5 suggests that $r_{e_t.e_{t-1}} = 0$ and applies only to time series data. If not satisfied, autocorrelation exists, resulting in incorrect estimates of both parameter values and their variances. Most critically of all, assumption 8 suggests that $r_{\times 1.\times 2} = 0$. If this condition is violated then multicollinearity exists, which, where the incidence is serious, may again result in parameter estimates which are both biased and inconsistent.

The potential violation of assumption 7 is a problem in accounting research, and one which is frequently overlooked. This is particularly so when dealing with *latent* explanatory variables (i.e., those which cannot be measured directly but which are founded on multi-item measurement constructs). This would include such familiar variables in behavioural and organisational research as organisational effectiveness, job satisfaction, budgetary participation, etc. If the instrument used to measure the variables displays a Cronbach alpha of less than one (which it always will!), then we have measurement error; if, more realistically, we have an alpha of only 0.8, then this may be acceptable on other grounds, but the significance of the measurement error could lead to biased regression coefficients and inefficient tests of statistical significance. Shields and Shields (1998) suggest that multiple regression methods may be inappropriate in such circumstances and that structural equation modelling may provide a more suitable alternative.

After fitting a regression equation we must conduct at least three tests of the violation of ordinary least squares assumptions, all of which may provide evidence of the mis-specification of the OLS equation:

- Monitor the size of correlation coefficient between X variables.
- Confirm that the explanatory variables (X_i) are independent of the residuals (e_i), and the absence of heteroscedasticity by ensuring that the correlation coefficient $r_{e^2.x_i}$ is not statistically significant for any explanatory variable. Graphically, this may be apparent from a wedge-shaped X–Y scatter indicative of a size relationship so that e_i increases as X_i increases.
- Ensure that for time series data $r_{e_t\, e_{t-1}}$ is not statistically significant or, alternatively, that tabular values of the Durbin–Watson d-statistic are within acceptable bounds. Graphically, plots of e_t against e_{t-1} in successive time periods should be random, but they may reveal a positive relationship (through clearly increasing or decreasing trends) or a negative relationship (through a saw-tooth pattern). Both are indicative of key variables omitted from the regression equation.

In a multivariate situation, where we have more than one explanatory variable, we seek to improve on the explanatory power of the equation (R^2) while at the same time ensuring that:

- coefficients remain statistically significant;
- coefficients and standard errors remain relatively stable;
- signs of coefficients remain intuitively correct.

Unless there is convincing evidence to the contrary, we begin by fitting a linear relationship of the form

$$Y = a + bX_1 + cX_2 + dX_3 + \ldots$$

A forward or backward stepwise regression procedure can be employed for the purpose (e.g., SPSS-X), with additional variables appearing in the equation, as long as they add to the explanatory power of the equation and they are individually statistically significant. The resultant equation should:

- have the highest possible explanatory power (using the adjusted R^2 feature to filter out useless variables making a negligible contribution);
- have a combination of variables for which appropriate tests ensure that all variables in the set are statistically significant;
- demonstrate the non-violation of the assumptions implicit in the use of ordinary least squares regression.

Figure 4.12 details the typical output from statistical software for regression results. The data for the illustration is drawn from Smith (1997) and his charity shops case study, which relates retail turnover performance to store characteristics. The key features of the output are:

- The inclusion of the three explanatory variables (DÉCOR, helpers and population) which correlate well with turnover (R = 0.684).
- The R^2 of 0.468 indicates that 46.8% of the variation in turnover for this sample of 56 cases is explained by the three independent variables.
- The adjusted R-square = 0.438. R^2 is adjusted to reflect the goodness of fit of the model by scaling down the R-square value in accordance with the number of explanatory variables in the equation.
- Standard error = 17.984, which is the standard deviation of the actual values of Y (turnover) about the regression line of estimated Y values.
- Analysis of variance (ANOVA) shows that the complete set of regression coefficients (i.e., the equation as a whole) is statistically significant (F = 15.272 significant at the 0.000 level).
- The column headed B gives the regression coefficient for the equation and establishes the equation as:

FIGURE 4.12

MODEL SUMMARY

R	R-square	Adjusted R-square	Standard error of the estimate	Change statistics				
				R-square change	F Change	df1	df2	Sig. F change
.684	.468	.438	17.984	.468	15.272	3	52	.000

Predictors: (constant), HELPERS, POPULATION, DÉCOR

ANOVA

	Sum of squares	df	Mean square	F	Sig.
Regression	14819.081	3	4939.690	15.272	.000
Residual	16818.929	52	323.441		
TOTAL	31638.010	55			

Predictors: (constant), HELPERS, POPULATION, DÉCOR
Dependent Variable: TURNOVER

COEFFICIENTS

	Unstandardised coefficients		Standardised coefficients		
	B	Standard error	Beta	t	Sig.
(Constant)	8.946	6.6447		1.383	.172
DÉCOR	6.039	1.646	.407	3.668	.001
POPULATION	.112	.042	.269	2.651	.011
HELPERS	.920	.280	.363	3.282	.002

Dependent Variable: TURNOVER

Summary of regression results for charity shops case study

$$\text{TURNOVER} = 8.946 + (6.039) * \text{DÉCOR} + (0.112)* \text{POPULATION} + (0.920) * \text{HELPERS}$$

- The column headed Beta gives the regression coefficients in a standardised form, with a zero intercept term. Standard error of B is a measure of the sampling variability of each regression coefficient.
- Column t measures the statistical significance of each of the regression coefficients by computing the ratio of (B/standard error of B) for each variable. Its level of significance is also displayed.

Moderated regression analysis (MRA)

Whereas the normal multiple regression equation looks like:

$$Y = a_0 + b_1X_1 + c_2X_2 + e$$

that for moderated regression analysis looks like:

$$Y = a_0 + b_1X_1 + c_2X_2 + d_3\ (X_1 * X_2) + e$$

and contains an additional 'interaction' term. The product term $(X_1 * X_2)$ represents the 'moderating' effect of X_2 on the relationship between Y and X_1. That is the relationship between Y and X_1 is conditional on the value of X_2. The variables X_1 and X_2 represent the main effects, and $(X_1 * X_2)$ the two-way interaction effect.

Where the coefficient 'd' is statistically significant, then the variable X_2 has a significant moderating effect on the relationship between X_1 and Y (as correspondingly does X_1 on the relationship between X_2 and Y). Adding further explanatory variables to the equation facilitates the evaluation of three-way, and even four-way, interaction effects.

In the accounting literature, moderated regression analysis has been widely used in contingency research, most notably in association with budgetary behaviour (e.g., Brownell, 1982; Otley, 1980). Hartmann and Moers (1999) provide examples of this type of research together with a critique of the application of MRA methods.

Further problems may arise when the moderated regression equation involves latent (i.e., unobserved) explanatory variables. Interaction terms are then likely to include significant measurement errors, suggesting that this form of regression analysis may not be the most appropriate method of estimation in such circumstances.

Structural equation modelling

Bollen and Long (1993) regard structural equation modelling (SEM) as an umbrella classification that covers path analysis, partial least squares and latent variable SEM, each as a preferred alternative to OLS regression in prescribed circumstances:

- Traditional multiple regression is confined to a single dependent variable and a number of explanatory variables. Path analysis (see Pedhazur, 1982) provides a natural extension by facilitating an analysis of the interrelationships in the variables so that the dependent variable from one equation can become the explanatory variable in a second equation. However, Maruyama (1998) observes a problem common to both path analysis and conventional regression methods in that only one direction of causation is permissible. Chong and Chong (1997) provide an example of the application of the use of path analysis in the accounting literature.
- Partial least squares (PLS) may be viewed as the 'poor man's' SEM in that it is the alternative sought when we cannot satisfy the stringent

assumptions of SEM. PLS may be preferred when we have a weak theory, small sample sizes (less than 100), and data which are likely to violate assumptions of normality. The weakness of theory often leads to PLS being referred to as 'soft modelling', being used for predictive rather than explanatory purposes. PLS attracts further criticism on technical grounds in that its construction of *latent* variables means that they are not latent in the conventional sense, but merely weighted linear additive combinations of observed variables. There are few instances of the use of PLS in the accounting literature, but Ittner, et al. (1997) provide a rare example.

The development of SEM follows Anderson and Rubin's (1949) initial discussion of the use of maximum likelihood methods in the estimation of parameters for single equations which form a complete set of stochastic equations. There is general acceptance (e.g., Hair et al., 1995; Kline, 1998; Schumacker and Lomax, 1996) of the two-stage technique suggested by Jöreskog (1969) for explaining the relationships between latent variables:

1. A measurement model which uses confirmatory factor analysis of covariance structures to demonstrate the relationships between the observed variables and the constructs (latent variables).
2. The development and estimation of a structural equation model, which is the representation of a causal path diagram through a set of linear equation.

There is no single measure of the goodness of fit of structural equation models that is acceptable under all circumstances (i.e., method of estimation, sample size, software employed). Accordingly, most published papers will report multiple-fit measures; this extends to the reporting of eight different fit indicators in Fogarty et al. (2000)! Smith and Langfield- Smith (2002) note the importance of reporting the specific modelling software employed since this will likely impact on the relationships explored. Thus the AMOS software (see Byrne, 2001) would not permit some of the relationships adopted by Collins et al. (1995) when using LISREL, on the grounds of logic. SEM software permits the adoption of a 'model-generating approach', which suggests additional paths to be evaluated, even though these may not have been specified initially on theoretical grounds. Such a facility has lead to widespread accusations of 'data mining' in the use of this approach; insignificant paths will usually be eliminated unless their removal results in a drastic reduction in the goodness of fit of the model. Thus Jaworski and Young (1992), Smith et al. (1993) and Baines and Langfield-Smith (2001) all add and drop paths to or from their initial models to improve the model fit, where such procedures still preserve theory and logic.

Smith and Langfield-Smith (2002) report the results of a survey of the use of SEM in the accounting literature, noting the increasing penetration

of the technique, though it remains at low levels relative to that of other behavioural disciplines.

Discriminant analysis

The use of ordinary least squares regression methods in the previous section requires a dependent variable that can be measured continuously. However, there will be occasions where the variable that we want to explain and predict is not of a continuous nature. It may be categorical of the form high/medium/low, good/bad, or success/fail. These can be quantified by assigning dummy variables of the (1, 2, 3) or (0, 1) variety to reflect the alternative states, but in each case these are the only values that the dependent variable can take. Changes in the value of the explanatory variable cannot change the continuous value, only its classification into one or other of the categories. In such circumstances we cannot use simple regression methods but must seek an alternative. Linear discriminant analysis (LDA) can be used when:

- the groups being identified are clearly separate;
- the explanatory variables are close to being normally distributed, or can be transformed to be so. This ensures 'univariate normality', where the stricter requirement of 'multivariate normality' is more difficult to test in practice;
- there is no multicollinearity between the explanatory variables.

We seek to construct an equation of the form:

$$Z = a + bX_1 + cX_2 + dX_3 + \ldots$$

such that the resulting value of Z allows the categorisation of cases. Effectively, we are generating the equation of a line (or lines), which can be positioned to divide the cases into the required groups. For example, in a failure prediction scenario, the construction of a three-variable discriminant model using financial ratios representing profit, debt and liquidity, may be visualised relative to the space in a rectangular room where axes are constructed in the corner of the room: the profit ratio stretches vertically towards the ceiling, and liquidity and debt axes are at right angles along the skirting boards. The company cases under consideration appear as points in space, representing three-ratio combinations, and discriminant analysis would try to position a plane in this space such that all the failed companies were on one side of this plane and all the healthy ones on the other. The equation of the optimum plane, even if it were impossible to classify all company cases correctly on either side, would be given by:

$$Z = a + (b * \text{Profit}) + (c * \text{Liquidity}) - (d * \text{Debt})$$

where

b, c and d are the weighting attached to each of the ratios

a is a constant term whose value determines the cut-off between failed and non-failed groups, and

Z is the value of the composite function, such that

$Z > 0$ corresponds with a state of financial health and

$Z < 0$ corresponds with a state of financial distress, in that the company has a financial profile similar to that of a previously failed company.

The exhibition of a negative score does not necessarily foreshadow bankruptcy, but gives an indication of financial distress in that the company has the profile of a previously failed company. The negative score therefore provides early warning in that future failures will almost certainly come from this distressed group, members of which require close attention.

Multivariate techniques such as linear discriminant analysis (e.g., Altman, 1968; Altman et al., 1977; Taffler, 1983) or LOGIT (e.g., Ohlson, 1980; Zavgren, 1985) have been used to generate appropriate models that best discriminate between samples of failed and non-failed firms based on a set of computed financial ratios. The derived model can then be used to classify (predict) other firms as potential failures or as financially healthy. Typically, such models are able to distinguish between failed and non-failed firms with very high degrees of accuracy (e.g., Altman, 1993, pp. 219–20; Taffler, 1995), and are widely used in practice (e.g., Altman, 1993, pp. 218–19; Taffler, 1995).

Smith and Taffler (2000) detail the classification of failed and non-failed companies on the basis of their narrative content within the context of two-group linear discriminant analysis (LDA). Their discriminant function is of the form:

$$Z = d_0 + d_1v_1 + d_2v_2 + d_3v_3 + \ldots$$

where Z is the discriminant score, $\{v_j\}$ are the variables selected for inclusion in the analysis and $\{d_j\}$ are the optimal coefficients, with d_0, the constant term, representing the cut-off criterion between the two groups.

To take account of the differential costs of Type 1 and Type 2 errors, C_1 and C_2, that is, classifying a failed firm as non-failed and vice versa, and the differential proportions of potential failures and solvent firms in the corporate population, p_1 and p_2, d_0 is adjusted according to

$$\ln\left(\frac{P_1}{P_2} \cdot \frac{C_1}{C_2}\right).$$

Altman (1993, pp. 254–63) provides empirical evidence that for the commercial bank loan decision in the USA the ratio of p_1/p_2 is 2/98 and the C_1/C_2 ratio is around 31 times.

Whereas alternative multivariate methodologies such as quadratic discriminant analysis (Altman et al., 1977), LOGIT and PROBIT (e.g., Ohlson, 1980; Zavgren, 1985), non-parametric methods such as recursive partitioning (Frydman et al., 1985), and neural nets (e.g., Altman et al., 1994) are detailed in the literature, there is no evidence of significantly superior performance associated with such approaches compared with traditional LDA (e.g., see Hamer, 1983; Lo, 1986). This is probably because the classical linear discriminant model is quite robust in practice (e.g., Bayne et al., 1983).

Ideally, the validity of a classification model needs to be tested by seeing how well it predicts (classifies) other cases in a 'hold-out' sample. If a model fails to perform well for the hold-out sample, two possibilities exist: (1) the model is sample-specific, or (2) the hold-out sample is not representative of the population from which it was drawn. Since data are difficult to collect, especially matched samples, hold-out samples are often seen as a luxury. A common solution to the problem is to test the validity of the model by varying the cases on which the model is based.

Validation of the derived models is undertaken using the Lachenbruch (1967) jackknife hold-out test approach, which provides almost unbiased estimates of the true misclassification probabilities. In this approach, $(n_1 + n_2)$ discriminant functions are computed from the original data samples of size n_1 and n_2 observations, with a different observation held out each time, which is then reclassified by the function computed from the remaining $(n_1 + n_2 - 1)$ cases. If m_1 and m_2 observations, respectively, are misclassified in the two groups, then the ratios m_1/n_1 and m_2/n_2 will provide the almost unbiased estimates of the true misclassification probabilities.

Simnett and Trotman (1992) identify three key reasons for non-use of financial distress models in practice:

- lack of formal training of most practitioners;
- criticism of the statistical assumptions underlying the models;
- failure to include non-financial variables widely accepted as useful discriminators.

More research is necessary to address the third of these issues, but it is possible to demonstrate the usefulness of such models in the appraisal of financial performance, while recognising the limitations to their implementation.

LOGIT and PROBIT

Discriminant analysis may not be the most appropriate method of estimation in some circumstances because the data violate key assumptions of its application:

- The explanatory variables are assumed to have a multivariate normal distribution. However, many financial variables – notably financial ratios used in failure prediction – are not normally distributed (Eisenbeis, 1977;

McLeay, 1986). Variables which have a lower bound of zero are inevitably non-normal.
- Samples are assumed to have been drawn randomly from their respective populations. These conditions are rarely satisfied in failure prediction studies because of the rarity of the failure event and the likelihood that a random sample would generate few, if any, failed cases. The widely used matched-pairs technique for case selection specifically violates this assumption.
- Group covariance matrices should be equal if a linear classification method is to be used – an assumption which is rarely satisfied.

Logistic (LOGIT) and probabilistic (PROBIT) regressions do not necessitate such restrictive assumptions, and are thus often more appropriate estimation methods, even though, as we saw earlier, the predictive outcomes may be little affected. Importantly, sampling techniques, which do not involve case-matching, permit the inclusion of variables like size, industry and economic cycle in the analysis of business failure. A number of researchers (e.g., Koh, 1991; Lennox, 1999; Ohlson, 1980; Zmijewski, 1984) have adopted LOGIT and PROBIT methods in the accounting literature. The logistic regression model estimates the model:

$$\text{Log } [P_i/(1 - P_i)] = a + B_1X_1 + B_2X_2 + \ldots B_kX_k$$

where P_i is the probability of the event under study.

LeClere (2000) observes that the elasticities of probability provide the best indication of the effect of independent variables on the probability of bankruptcy. Following Ohlson (1980), Shumway (2001) uses logistic regression to model financial distress, on the basis of the probability of subsequent bankruptcy as:

$$\text{LN (odds ratio)} = -7.811 + (4.068)X_1 - (6.307)X_2 - (0.158)X_3 + (0.307)X_4$$

where X_1 = TL/TA
X_2 = NI/TA
X_3 = CA/CL
X_4 = LN (AGE)

The coefficient of –6.307 for the NI/TA (net income variable) allows the exponent to be interpreted: $e^{-6.307} = .0018$, so that a one unit increase in NI/TA decreases the odds of bankruptcy by 0.18%, when the other covariates of the model are controlled.

In other than failure prediction environments, Trubik and Smith (2000) use logistic regression to develop a model which explains customer defection in retail banking. They report a four variable model to examine the odds of staying versus leaving:

$$\text{LN (odds ratio)} = -7.9439 + (0.7031)X_1 + (3.8607)X_2 + (2.4333)X_3 + (0.9933)X_4$$

where X_1 = time in years an account has been held with the bank
X_2 = fee exemption levels
X_3 = number of bank products held
X_4 = delivery channels used

As with linear discriminant analysis, the cut-off for classification purposes is determined by the relative costs of misclassification. Thus for failure prediction modelling we might be guided by the relative cost of Type 1: Type 2 misclassification errors. In our customer defection example, the literature suggests a 10:1 cost comparison between attracting a new customer and retaining an existing customer, and a cut-off value of 0.05, which correctly classifies 83.2% of bank customers.

Multivariate Analysis of Variance (MANOVA)

If ANOVA were employed consecutively for dependent variables that were eventually shown to be interrelated, then unreliable inferences would result. In such circumstances it is better to use a method (MANOVA) which simultaneously tests all of the variables and their interrelationships. MANOVA is thus very similar to ANOVA and operates on the same basic principles, except that it can handle more than one dependent variable. As with ANOVA, MANOVA uses the F-test to measure the differences between groups by comparing within-group variance to between-group variance.

MANOVA evaluates the differences between the multivariate means (centroids) of several populations, on the null hypothesis of equality. When the null hypothesis can be rejected, follow-up tests can be performed to pinpoint the source of the difference, namely:

- univariate F-tests run on each of the dependent variables; or
- multiple discriminant analysis.

The dependent variables for use in such an analysis must be correlated with each other, otherwise there is no justification for using MANOVA – ANOVA with separate F-tests would then be more appropriate.

The content of this chapter is inevitably a compromise, both in terms of the techniques addressed and the depth of coverage attempted. A search of the recent accounting literature, especially those journals favouring a quantitative approach, will reveal more complex methods and tests than are described here. However, those considered above still represent the analytical tools most likely to be employed and reported. The mathematics of quantitative methods can be intimidating, and can cause some researchers to neglect the

potential benefits of carrying out even the most cursory of tests of hypothesis. However, the most important elements of this chapter are, arguably, the least quantitative:

- the recognition of a theory which allows us to establish 'expected' frequencies and outcomes, with which to compare our observations;
- the basic principles of sample selection; and
- measurement issues.

The number-crunching (of ever increasing sophistication) to handle increasingly complex experimental designs is secondary, building on these foundations to demonstrate the potential which exists for significance testing and model building.

5

Research Ethics in Accounting

CHAPTER CONTENTS

Some of the atrocities committed during the Second World War in the name of medical research prompted the development of a generally accepted code of research ethics which has spread from the medical arena to all other research disciplines. Unfortunately, the transfer of such codes to the business discipline has often been seen by researchers to be an imposition, unnecessarily constraining their activities, and something better left in medicine. Such attitudes display an ignorance of the fundamental objectives of ethical guidelines and the benefits that they can convey to all concerned.

The best-documented abuses by misguided researchers remain in the medical field. Thus Dooley (1995, p. 24) reports on the Tuskegee Syphilis study in the USA, where researchers conducted a 40-year longitudinal study, right up to the 1970s, of an isolated black community with high incidence of the disease. Even though effective treatment through penicillin was widely available, the researchers actively sought to prevent subjects from being treated, and to avoid the disclosure of deaths caused by untreated disease, because of potential distortion to their research findings.

This may be an extreme instance, but the potential for the researcher to act in a manner which is not in the interests of the participants exists in all disciplines, especially where the participant has little control over the events associated with the research. For example, in the accounting education sphere, it may be possible to devise a field experiment where two groups of students were taught in parallel using very different methods in order to evaluate the relative superiority of techniques. But if this experiment were over an extended period (say ten weeks), and after week two it became clear that one method was vastly superior to the other, could we on ethical grounds

continue the experiment, when one group of students was clearly being disadvantaged? The natural focus would therefore be on business scenarios in which participants could be exploited in a manner that endangers their future welfare, but ethical considerations are much wider than this in practice, and concern issues of honesty, trust and subjugation.

Australia's National Health and Medical Research Council specifies the scope of research ethics in terms of three underlining principles: 'the protection of the welfare and the rights of participants in research; to facilitate research that is or will be of benefit to the researcher's community or to humankind; to provide a national reference point for ethical consideration relevant to all research involving humans' (NHMRC, 2000, p. 3).

The ethics quiz

The following 'ethics quiz', inspired by Agnew and Pyke (1994, p. 273), aims to explore some of the confusion and misunderstanding that surrounds ethical issues in accounting research. You have to determine whether the key figure in each of the following scenarios has behaved ethically or not.

1. Professor A has sought the collaboration of the CEO of DeltaCorp in the conduct of interviews with the workforce. The CEO has agreed, on the condition that no further separate permissions are sought from the workers; he has provided permission and the workforce will participate as part of its job. Professor A agrees to continue on this basis.
2. Candidate B is conducting interviews with accounting managers in Hong Kong. She suggests that the requirement to have interviewees 'sign-off' on completion of the interview is an insult to the participants, which will cause loss of face. Her supervisor, Dr C, agrees that the ethical guidelines need not be observed in this instance because of cultural differences.
3. Associate Professor D is applying for promotion and, in order to strengthen her case, has sought testimonials from graduate students under her control as to the quality of her teaching, research and supervision.
4. Professor E supervises a number of Masters and PhD students. Contact typically involves discussions prior to the commencement of the study (usually not exceeding five hours in total), some discussion concerning analysis of the results, and the review of one or two thesis drafts. Professor E insists on co-publication of any research papers, where he is invariably the first-named author.
5. Associate Professor F used deception in her study of the impact of gender and self-esteem on accounting decision-making. Prior to participation, all subjects are informed of the requirements and purpose of the experiment as much as possible, given the deception component, and

of their right of withdrawal. Following an initial accounting task, an assessment of gender-role orientation was conducted: all female subjects, irrespective of actual performance, were told they had exhibited a masculine orientation; all masculine subjects that they had revealed a feminine orientation. Measures of self-esteem and a second accounting task were then administered. On completion, subjects were thanked for their help and promised a detailed report of the outcomes of the study. Two months later subjects received this report, which fully described the deception.

6. Dr G is conducting a field study examining the relationships between performance and management control systems at a number of different locations of the same organisation. There is a friendly rivalry between the locations with each keen to be seen to be outperforming the others. Dr G does not disclose any performance or budgetary information, but he is willing to talk about differences in organisational structure and management style.
7. The Director of an Accounting Research Centre is aware that one of his most consistently successful grant-winners is behaving in a seriously unethical manner. Despite repeated warnings about violation of ethical standards, the colleague's behaviour persists. The Director decides to take no further action.
8. A requirement for Dr H's computer-based accounting course is participation in an extended multiple-choice experiment. Minor electric shocks are to be administered to the fingers of students who make an error, in the belief that it will improve their subsequent performance. On learning of the nature of the experiment, one student seeks to withdraw, but his protests are waived aside by Dr H, who insists that participation is a course requirement.
9. Professor J is leading a team of researchers examining the achievement, motivation, creativity, personality and numeracy of Year 10 school students to determine their suitability for enrolment in an accounting undergraduate course. The Principals of some of the schools have requested copies of all the test scores for each of their students by name. The researchers provided the information in the interests of preserving the continuing goodwill of participating schools.
10. Dr K has published the results of a large study examining the response of investment analysts to changing levels of accounting information disclosure. She subsequently receives a request from the Stock Exchange for access to her data so that they can re-analyse the data and confirm her conclusions. Dr K refuses the request on the grounds that a computer crash has caused the loss of part of the dataset.

Each of these practices is likely to be in breach of typical ethical guidelines for accounting research because they fail, in one way or another, to observe acceptable relationships with human participants.

1. The principle of 'informed consent' is fundamental to the conduct of ethical research. It appears that the employees of DeltaCorp will have no opportunity to withdraw their participation, without threatening their continuing employment. Professor A should not continue to pursue the research project on this basis.
2. The transplanting of 'Western' guidelines to other cultures will inevitably cause difficulties. Where the data collection is being conducted offshore for a research degree in another country, then the ethical guidelines of the host university should be observed, even though their implementation causes practical difficulties.
3. Testimonials should not be solicited from persons who, because of their particular circumstances (in this case as part of a close supervisor–supervisee relationship), are vulnerable to undue influence.
4. Publication credit should reflect the contributions of the parties involved. Unless Professor E made a significant contribution to the final version of the published paper, he should not receive co-authorship. Where a paper is based on a thesis, the author of the thesis would normally expect to be first author.
5. Although Professor F was sensitive to certain ethical issues, there was no attempt to detect and remove any potentially damaging consequences for the individual participants arising from the deception. Any anger or resentment arising from the eventual disclosure of the deception was apparently neither monitored nor evaluated.
6. Dr G has contravened his obligation to safeguard the confidentiality of the information obtained from the different locations of the organisation during the course of the research.
7. The Director has failed in his duty to bring these unethical activities to the attention of the appropriate committees on ethical standards and practices for the accounting profession.
8. The investigator must respect the individual's freedom to decline to participate, or withdraw from, research at any time. The investigator should have provided the student in this case with a choice of alternative activities to fulfil course requirements. The research may also breach requirements associated with protecting the subjects from physical discomfort.
9. This form of reporting is unethical since there is no indication of any limitations on the information provided. There has apparently been no attempt to examine the potential misuse of the information supplied.
10. Once research results are published, the data should not be withheld from other competent professionals as long as the confidentiality of the participants is maintained and protected. The researcher is obligated to keep the data safely so that verifications of this nature can be carried out.

Although the scenarios above are hypothetical, they are close to many of the situations that arise in practice and that will cause concern to researchers, especially where they appear to be faced with a choice which means that they

either compromise their ethical principles or sacrifice the outcomes of a research effort.

Informed consent

Strict guidelines, based on the medical research model, are applied, and adherence is monitored in business research too, extending to written consent of respondents and a code of voluntary participation. Thus any research involving human participants would require approval of a university ethics committee. This would embrace research involving data collection by interviews, questionnaires, focus groups and observation. Indeed, informal professional conversations with practitioners, while not normally requiring ethics approval, would do so if such conversations were systematically employed by the researcher as a means of collecting data. Even if there is no human participation, archival research which leads to accessing medical records, or other sources which are not publicly available, and contain sensitive information (e.g., identifying people by name) still requires ethics approval. Where the agreement of other organisations is necessary for the research to be conducted, then written organisational approval (from the CEO or similar) to access staff members, clients or proprietary information is necessary. This would extend, for example, to written permissions from Deans and/or Heads of School, as appropriate, to access students for research purposes – even though no harm will result from their participation. Informed consent and anonymity are paramount in the process.

Some of these situations can be extremely problematical. Even gaining access to a site in order to conduct research is becoming increasingly difficult. Firms often need to be convinced that there is 'something in it for them' before granting permission for research to be conducted. While researchers need to be able to demonstrate such benefits, they must also be aware of the potential disadvantages for other participants that can arise. A number of examples from my experience are illustrative:

- A firm employs a research student to conduct a benchmarking exercise as part of a research degree. The student is asked to survey competitors to determine their current practices, but not to reveal their affiliation to the employing firm. This situation is one which could easily be represented as 'industrial espionage', especially where the base firm is unscrupulous in its use of the information gathered covertly.
- A firm grants permission for the conduct of staff interviews, with the view that it becomes 'part of their job'. The boss has given permission and they will accede to his/her wishes. A researcher properly seeking the informed consent of the participants may be faced with both reluctant subjects (who feel they have to take part, even unwillingly) and the wrath of a boss who

is not used to having his/her authority questioned (and who may consequently refuse any further collaboration).
- A firm willingly engages in research concerned with the implementation of new business processes, but its participation is part of a wider downsizing agenda. As a result, some of the subjects involved in the research may subsequently lose their jobs indirectly because of the responses they have made as part of the research.
- In order to gain access in the first place, Hammersley and Atkinson (1983) highlight the importance of the initial impressions created by researchers and the impact that these may make on the successful completion of research projects. Such impressions may secure comprehensive access, or cause access to be restricted or denied completely. They suggest that 'impression management' with respect to such issues as dress, gender, age, attractiveness, ethnicity, manner of speech and perceived level of expertise may all facilitate the conduct of the project. The creation of a favourable impression therefore contributes to the success of the process, but at what cost to ethical considerations? Such findings have implications for the use of minorities and for changing acceptable stereotypes, since to secure access we may have to provide 'acceptable' researchers, in a similar manner to the audit client requirements observed by Grey (1996), which restricted the opportunities for young female auditors.

Several of the preceding chapters, detailing the implementation of specific research methods, highlight particular ethical concerns. Those relating to complete honesty, and the disclosure of the whole truth, for example, generate a number of concerns:

- The use of experimental methods identifies instances where completely honest disclosure is avoided in order to preserve the integrity of research outcomes. Smith and Taffler (1996) report the use of, arguably harmless, untruths in setting the task requirements in judgement decisions – telling participants that they are evaluating separate companies even where there is considerable replication – in order to induce unique decisions. Trotman (1996) emphasises that the researchers must make the purposes of the research clear to participants, but not in such a way that they might deduce the research hypotheses, since this would impact on the internal validity of the study. However, Gibbins (1992) warns that if research subjects have been deceived by researchers, this will alter the way in which they receive future requests to participate. Without straining the strict 'letter of the law' of ethical guidelines in terms of deception, some research, particularly experimental studies, cannot be contemplated.
- Involvement in field studies offers the possibility of covert research. Where researchers are involved in studies where other participants are unaware of their role and objectives, then both honesty and trust must be jeopardised. However, without stretching the ethical guidelines in this way, it is doubtful whether potentially important findings would emerge.

- The imposition of ethical requirements, and the subsequent interaction between ethics committee and candidate, may strain the traditional supervisor/candidate relationship, especially where this may be construed as interference in the implementation of selected research methods.

Ethical guidelines

Ethical issues in business research extend from those concerned with the conduct of the research through to the publications process subsequent to the research. Some of these issues remain hopelessly under-addressed in many universities (those in the UK are a good example) especially compared to what is undertaken in Australian universities. An apparent conflict with the traditional all-embracing role of the supervisor may contribute to the problem, especially where ethics committees are perceived to be interfering unnecessarily.

However, some of the apparently acceptable US practices (e.g., those on co-authorship, see Coppage and Baxendale, 2001) may be construed to constitute academic malpractice in Australia. Typically, co-authorship of a paper would involve substantial participation from all authors in its construction, embracing *all* of the following conditions:

- conception and design, or analysis and interpretation of data;
- drafting the paper or revising it critically for important intellectual content; and
- final approval of the version to be published.

Clearly supervision or funding of the work are not grounds alone for appending one's name to the work of a student!

Any consideration of the ethics of a research proposal would normally address at least the following issues:

- appropriate written permissions from participating organisations;
- eliminating opportunities for personal harm, physical or mental, to research participants, including the researcher;
- informing participants of the motives for the research;
- providing feedback of the results to the participants;
- gaining permission from participating individuals (other than for mail surveys, where return of the questionnaire is taken to imply permission);
- avoiding coercion in management settings;
- guaranteeing and delivering both confidentiality and anonymity to the participants;
- granting the right of withdrawal to participants at any time;
- guaranteeing the safe storage of research data, usually for a period up to seven years.

As we will see, it is not always possible to satisfy all these conditions absolutely and still ensure the integrity of the research approach. Hartmann (2000) emphasises three issues which cause the greatest difficulty for researchers in securing ethics approval from their university scrutinising bodies:

- A clear view of how the research results will be used, especially in situations where the research findings will be made available to the management of the host organisation, but which may be of restricted availability to the participants themselves.
- The issue of consent, especially in action research projects. The circumstances of participation require a level of involvement, honesty and openness of communication unusual for both the organisation and the researcher. Hartmann observes that 'guarantees by the researcher of confidentiality may be meaningless in the long term; such research is common and yet current guidelines mean that it cannot occur, or the usual consent procedures are meaningless' (Hartmann, 2000, p. 6).
- The importance of issues associated with national and organisational culture complicates the behaviour of both individuals and groups, and constrains their actions in a manner that may invalidate the research outcomes. Some flexibility has to be imported to cope with cultural differences. For example, the requirement for written consent of individual interview participants, subsequent to the interview, causes considerable problems associated with 'loss of face' in South East Asia. Frequently, tight guidelines have to be relaxed because a formal 'signing off' is often seen to be insulting to those involved and too great an imposition!

We have a professional duty as academics to inform both our student and practitioner audiences of the outcomes of current research and their implications for practice. Students should be placed in a position where they are able to question the whole research process by providing critical comment on alternative methods. For example, the running of quasi-experiments in class with the express intention of exposing their limitations rather than collecting data can be most rewarding, clarifying exactly what can go wrong while increasing levels of cynicism all round. Similarly, reviewing the survey instruments on which past publications have been based helps to expose both the ambiguity and uncertainty of this particular research process so that nonsensical outcomes become all the more understandable, though none the less palatable.

Honest and transparent reporting of research practice is an ethical duty of those participating in accounting research. Researchers should report everything that they did, why they chose that course of action, and how the procedures were conducted. Any doubts that are apparent at any stage of the research should be highlighted, along with their implications and the actions taken to overcome deficiencies, in the stated limitations to the research. Where researchers appear to have been 'economical with the truth' in their

reporting, this is normally apparent in their papers and is indicative of poor research.

Hartmann and Moers (1999) report findings which are of ethical concern in respect of the availability of research data to other researchers. Despite published papers advertising the availability of data from a specified author, their attempts to obtain such data for re-analysis were thwarted. Their requests either went unacknowledged or met with excuses ranging from 'data lost in computer crash' through to 'data lost in move to a new university'. Such responses are hardly consistent with our ethical responsibility to guarantee the safekeeping of research data for at least seven years.

6

Experimental Research

CHAPTER CONTENTS

Abdel-Khalik and Ajinkya (1979) provide a precise definition of the nature of an experiment in that the researcher manipulates one or more variables with subjects who are assigned randomly to various groups. These groups receive different combinations of the variables (termed treatments); in some cases a control group may exist which receives no such treatments. The major advantage of experiments lies in the researchers' ability to ensure high internal validity, defined in terms of how well they can eliminate rival explanations for their results. Experiments are, thus, particularly suited to research questions that investigate causal relations between variables.

Gibbins and Salterio (1996, p. 24) suggest four guidelines for good experimental research in accounting:

1. A clear statement of the problem, its importance and the contribution to knowledge that its solution will make.
2. A clear statement of the theory that underlies the process, in particular the theory that drives the behaviour and the impact of context on theory.
3. A sound experimental design – since fatal flaws can be introduced by inappropriate or inadequate designs.
4. Recognition of the importance of external validity. If this means that we need more realistic experimental settings, then we need correspondingly richer theoretical explanations of resultant behaviour.

These guidelines provide an excellent framework for the consideration of research issues in accounting experimentation, and are adopted here.

The problem statement

The variables of interest are:

Y = the dependent variable or observation
Xi = the independent variable (or treatment) which can be either manipulated in value, as in an experiment, or measured from archival data, as in econometric studies. A laboratory experiment will normally have more internal validity than other research methods (particularly field studies or archival research). As we move away from the high-control environment of an experimental setting, the various threats to internal validity grow, with the consequence that tests of causality cease to be so reliable. We may then only be able to attribute association (through correlation measures), rather than the direction of the association.

Theory and context

With the explosion of experimental research in auditing in the mid-1970s, the use of experienced subjects or expert practitioners as participants was thought to be essential to provide the necessary external validity to the research. As a result, most journals have expressed a preference for the use of practitioners in any experimental research. This has imposed a serious constraint on the implementation of experimental work, particularly over the last ten years, because it has become increasingly difficult to secure the participation of Big 4 practitioners, say, for experimental research. This participation remains essential for the publication of auditing research in the top journals; but fortunately, for experimental work outside the auditing sphere there is increasing recognition that surrogate participants may have the necessary skill base to participate. The use of a large proportion of students in a study has been justified by a number of authors: (1) Ashton and Kramer (1980), in a review of research in both business and psychology, report that the information processing behaviour of students and 'real-world' decision-makers does not differ; (2) Abdolmohammadi and Wright (1987) argue that for highly structured decision tasks the performance of a student should not differ significantly from that of real-world decision-makers; (3) past research

has reported that there are no differences between subjects with and without work experience in experiments using financial information datasets (MacKay and Villarreal, 1987; Stock and Watson, 1984). Brownell (1995, p. 83) suggests that the alleged shortcomings of using students in laboratory-based experiments are overstated. He likens the situation to that of non-response in survey work, suggesting that we need to demonstrate both that systematic differences do exist and that such differences matter, before we rule out the use of student surrogates.

The role of theory in real judgement settings

In order to take advantage of the potential strength of experimental settings in testing causal relations, a well-developed theoretical framework is essential. Once the theoretical framework has been established, the researcher can decide which variables should be manipulated and measured, and which controlled.

Early experimental research in accounting was highly criticised for its atheoretical approach, but more recently we have seen the application of psychological theories to decision-making research, following the work of Kahnemann and Tversky (1972) in the exploration of heuristics and biases (e.g., Ashton, 1983; Smith, 1993) and more recently in the use of theories of knowledge, memory and learning (e.g., Libby and Frederick, 1990). Economics-based theories (notably agency theory) have also evolved to explain experimental settings in management accounting (e.g., Demski and Feltham, 1976).

The embedding of judgement within the context of a particular task

The question of context is a potentially great threat to the external validity of judgement in experimental tasks. We may be unsure whether the observed findings are attributable to the variables subject to manipulation, or to the impact of such variables as task complexity, location, time constraints, incentives, etc. A well-designed experiment should attempt to address each of these issues, but even so it remains difficult to control for all potential extraneous variables successfully with 100 per cent certainty in any experiment. Great care must be taken in assembling a literature of comparative findings to support a particular research direction, especially without returning to the original papers; inconsistent or contradictory findings may be attributable to different research contexts rather than to the key variables at issue.

The role of incentives to participants

One contextual issue is of particular concern and has raised a number of associated issues: the availability of incentives to motivate participants to expend effort in the performance of the task may create more problems than

it solves. Libby and Lipe (1992) provide an interesting example of this where they suggest that the conceptual noise introduced by the incentive scheme is influenced by unmeasured factors. For student participants, modest monetary incentives might be necessary to induce attendance, participation and attention, but the impact of these on the eventual outcomes is unclear. We may need to examine a theory of incentives, and the way they influence individuals with differing personal characteristics, before we can measure their effect. Libby and Luft (1993) suggest that any extra effort induced by incentives may be related to ability and knowledge characteristics of the participants. Ethical requirements may impose a further constraint in that while incentives to participate may be acceptable, performance-related incentives (where differential rewards are provided) may be disallowed on equity grounds. Hence, we may reasonably try to avoid the availability of incentives wherever possible. The use of incentives appears to be unnecessary where the participants are skilled professionals. Bonner and Sprinkle (2002) have recently established a theoretical framework for incentive payments, which is likely to generate further research in this area.

The use of professionals as research participants

As discussed above, some journals and some experimental settings demand the use of skilled professional participants. It is becoming increasingly difficult, even outside the top-tier journals, to publish experimental research using wholly student audiences. This is particularly so for undergraduate audiences. Student audiences may still be acceptable for the conduct of pilot studies, but practitioner/professional subjects will be required for the main study.

The use of deception to create appropriate research settings

The mere fact that subjects are placed in a laboratory setting may create an effect resulting in an outcome which would not have arisen outside the experimental setting. This can occur for a variety of reasons, including where the subjects may adopt an attitude whereby they try to please the experimenter: Weber and Cook (1972) and Schepanski et al. (1992) report on situations where the participants in experiments want to help the experimenter by apparently trying to deliver what is expected of them.

Under such circumstances it may be necessary to deceive the subjects as to the objectives of the experiment, so that they are unaware of the predicted outcomes of the study, in order to deliver valid results. Thus Smith and Taffler (1996) inform their subjects that they are making judgements on 60 separate companies, to secure separate decisions for each of the 20 company cases being employed with three different treatments. Only on completion of the experiment is the deception revealed. Such behaviour, however, may generate cynicism among participants so that they are in such a state of

disbelief regarding experimental settings that it may preclude future participation. There are also ethical considerations in that participation in the experiment has been achieved in a potentially undesirable manner, with the result that researcher trust and honesty have been sacrificed in order to deliver experimental outcomes.

Experimental design

Trotman (1996), following Campbell and Stanley (1963) and Libby (1981), suggests a number of simple alternatives.

Post-test only control group design

Subjects are randomly assigned to treatments, that is to different levels of the independent (explanatory) variables. For two levels of the independent variable, X_1 and X_2 (one of which may be a control group with no treatment), two corresponding outcomes are observed, OO_1 and OO_2, say. These outcomes are measured after each of the subjects has received the treatment. A comparison of OO_1 and OO_2 will reveal the impact of the different treatments. The basic form of this design could easily be expanded to cover many other levels of the treatment. Joyce and Biddle (1981) provide an example of this type of design.

Pre-test/post-test control group design

Subjects are measured to see how they react to successive treatments, but this time a before-and-after experimental design is used which allows the research design to control for individual differences. A treatment X_1 is applied to a subject, producing an observed outcome OO_1; a further treatment X_2 (which may be something as simple as the provision of additional information) is then applied to the subject and a new outcome OO_2 observed, facilitating the comparison of the two observed outcomes.

Despite the advantages of the control of individual differences, the repeated use of the same individuals for successive treatments also causes potential problems: there may be a 'learning' effect as well as an 'order' effect. Heiman (1990) provides an example of this type of design.

Factorial design: between subjects

This involves the simultaneous variation of two or more treatments (explanatory variables) so that we can monitor their separate impacts on the dependent

variable, and any potential interactive effect between the explanatory variables (i.e., the extent of any conditional relationship such that treatment X_1, say, only has an effect for certain levels of the X_2 treatment).

The simplest 2×2 design manipulates two treatments across two levels to give four different possible combinations. Subjects are randomly assigned to each of these four cells, and each receives only one treatment. Trotman (1996, p. 19) identifies a number of distinct advantages for using a factorial design of this type:

1. The interaction effects can be evaluated. This is particularly important where there are competing alternative explanations for the observations. Brown and Solomon (1993) provide an example of this type of design in testing three competing theoretical explanations from the psychological literature.
2. Potentially confounding variables (e.g., gender, work experience) can be held constant within a cell so that their influence can be evaluated.
3. External validity can potentially be enhanced by findings which demonstrate similar effects across a number of subject characteristics.
4. Designs of this kind are more economical than the simpler post-test only designs in their use of subjects. Fewer subjects are required to conduct the same tests, which is an important consideration where practitioners willing to participate are difficult to locate.

Factorial design: within subjects

As with the simpler designs, we can also introduce a 'before-and-after' factorial design. Whereas in a 'between-subjects' design subjects receive only one treatment, in a 'within-subjects' design each subject receives all the successive treatments. This constitutes what is often called a 'repeated measures' design. Such designs can have even further advantages:

1. They require many fewer subjects.
2. The statistical power of the outcomes is potentially greater because of the controls implicitly introduced for individual characteristics, assuming that the characteristics of the individual have not changed significantly between successive treatments (see the maturation effect (page 109), regarding internal validity concerns).
3. They are particularly useful for examining differences in information use and the impact of different treatments in learning and training environments.

But there are also a number of potentially serious disadvantages associated with this research design, which have been identified by Brownell (1995, p. 11) and Trotman (1996, p. 30):

1. Demand effects, such that the subject is able to glean the details of the experiment (e.g., the hypotheses under test, or the number of failed companies in a sample) to such an extent that his/her behaviour becomes different from that which would be expected (under real-world conditions).
2. One of the major downsides is that subject variables cannot be used as explanatory variables in a within-subjects design because individual characteristics cannot be altered across treatments.
3. Practice effects, such that subjects' behaviour changes as the experiment proceeds – they may perform better because of skills learning and knowledge accumulation, or worse because of fatigue or waning enthusiasm.
4. Carry-over effects, such that the way a second treatment is considered is highly dependent on experience from the first one (e.g., in successive failure prediction tasks, the expectation, justified or otherwise, that the number of failed cases will be identical). This may mean that independent decision-making may be impossible without using counterbalancing measures associated with the re-ordering of the tasks. Even with re-ordering, Schepanski et al. (1992) suggest that there may still be a problem, which may be overcome by introducing 'filler' tasks and lengthening the experimental period. However, this is a solution which will itself pose further problems. Maturation effects may also be apparent here in that within-subjects designs must inevitably involve some time lag, namely that between successive treatments, which will make them more vulnerable to this kind of internal validity threat.
5. Statistical effects, such that the equality of variances assumption is violated in within-subjects analysis of variance (ANOVA), making the F-tests biased. The outcome may not be serious but should be acknowledged by performing alternative checks.
6. A cue salience effect, such that the number of variables under focus in a within-subject experiment will be fewer than in the corresponding between-subjects experiment. Schepanski et al. (1992) suggest that this may pose an external validity threat. A major constraint in most experiments is the amount of time available to subjects for participation in the experiment, and the variability in the time taken by individual subjects. All researchers are constrained to some extent by the number of variables they can consider given the number and amount of time the subjects are available.

There will almost always be other potentially influential variables that threaten internal validity, and these must be dealt with by exercising some form of control. A number of alternative procedures exist.

1. **Control groups:** to hold extraneous variables constant. A controlled environment is established by the researcher being present during the conduct of the experiment, in contrast to the non-controlled settings like those used by survey researchers. Where control groups are not feasible, we

may seek to control those variables not directly involved in the relationships under investigation.

2. **Randomisation:** so that the distribution of the variable is equally likely across all of the independent variables. Subjects will be assigned randomly to each of the treatment cells. Random assignment assumes that the same factors will influence each treatment group, and that each should contain an approximately equal mix of these factors. The larger the sample size, the more likely this is to be the case. Where we have a small sample, and very small cell sizes, we must doubt whether randomisation will effectively dull the impact of individual differences among subjects.
3. **Holding constant:** so that the variable has the same value across all values of the independent variable.
4. **Matching:** so that the distribution of the variable is common across each of the independent variables. Thus, for auditing experiments, we may match the participants based on their experience; in financial accounting, companies may be matched on size. Matching effectively precludes the study of the 'matched' characteristic. Where this is a problem we may seek to include matched variables (e.g., size) as explanatory variables. However, such inclusion is not a completely satisfactory solution because the matched variables cannot then be randomly assigned to experimental groups. Thus, if they are found to be influential, we cannot rule out an alternative hypothesis that some unknown variable (proxied by the included variable) is responsible for the relationship.
5. **Counterbalancing:** using all combinations of ordering of treatments will overcome the impact of any order effect, but will necessitate the preparation of many different versions of the experiment. Thus Trotman and Wright (1996), in an audit environment, had half the subjects processing internal control aspects, followed by going concern aspects, while the other half conducted the going concern part first, followed by the internal control. Similarly, Smith and Taffler (1995), in a financial accounting environment, address three different information processing formats (narratives [N], ratios [R], graphs [G]) which require six different instruments, one for each of the possible order combinations (NRG, NGR, RNG, RGN, GNR, GRN). However, counterbalancing may not solve all the problems of order effects. Smith and Taffler (1996) observe that for tasks of similar difficulty, the 'last' test performed, whichever it was, was often performed the quickest.
6. **Ignoring variables:** either intentionally (if the variable is considered to have an inconsequential impact) or unintentionally (if its impact has not been considered at all!).

Each of the alternatives has potential pitfalls. In archival studies, for example, it is usually impossible to hold constant a large number of variables, or to match on many others. Matching can create almost as many problems as it is

attempting to solve. For example, in accounting studies (e.g., Smith and Taffler, 1992) it is common to match companies on the basis of industry; some measure of size, and even financial year-end. Such a process may severely constrain the sample size achievable, and may effectively eliminate the consideration of the influence any of the matched variables (notably size) may have. Some variables can safely be ignored on the basis of prior literature, but many others may still have to be ignored even though we are uncertain of their potential impact. Thus it is common (e.g., Schulz, 1999) for variables associated with individual differences between participants to be ignored, other than those for age, gender and experience. This is despite the growing evidence for the influence of numerous other factors (e.g., cognitive style, tolerance of ambiguity) on behaviour in accounting environments. Some variables may be so difficult to measure that we seek to employ proxy variables instead – for example, the use of size measures to proxy for political influence. Failure to include potential variables may result in a selection bias at the data sampling stage.

Manipulation checks can be used to help ensure that subjects really understand both the instrumentation and what is required of them, and to monitor the course of the experiment (especially at the pilot stage, where one is feasible) to ensure that the different treatments are producing changes in the same direction as anticipated.

The validity trade-off

We can identify three distinct validity concerns in the conduct of experiments: construct validity, internal validity and external validity.

Threats to construct validity

Construct validity describes the extent to which abstract constructs are successfully operationalised. This definition embraces both the extent to which the constructs are measured reliably and the extent to which they provide measures which effectively capture the essence of the abstraction. We attempt to interpret theory through the development of abstract constructs, and then seek to operationalise these constructs with measurable variables. Where this is not possible we must substitute proxy variables, which are measurable but which may be less than perfect proxies. Imperfect proxies will mean flawed tests of hypotheses.

Nunnally (1978) describes three aspects of establishing construct validity:

1. Specify those variables which are both observable and related to the construct – this tells the researcher which items to measure and evaluate in the next step.
2. Determine the extent to which these observable variables are reliable measures of one or more constructs and the extent of interrelationships among measured items.
3. Determine the extent to which those measures of the constructs employed will produce predictable results.

Most of the attention in designing laboratory research is on the experimental treatment with typically little attention paid to the construct validity issues associated with the measured variables. Where there is any attention being paid, it is usually to the second of Nunnally's stages alone.

The consequences of errors here are potentially serious. For example, Brownell (1995, p. 112) reveals that in his earlier paper (Brownell, 1982) he used the widely accepted Milani (1975) instrument to measure budgetary participation. However, he perceives 'participation' to be a combination of both 'influence' and 'involvement', and is not convinced that the Milani instrument adequately addresses both dimensions. The instrument used may not be measuring the construct in a reliable way, but the alternative is to develop a new and completely untested instrument. Brownell is apparently prepared to trade construct validity for reliability in this instance.

Threats to internal validity

Cook and Campbell (1979) identify nine separate internal validity concerns.

1. **Maturation.** This concerns any impact associated with the passage of time. This would relate to changes to subjects between successive experiments, or even during a single experiment or interview. It would also relate to the characteristics of company cases, say, where growth or restructuring over time mean that we are no longer comparing like with like. Thus the researcher must be aware of the potential impact on subjects of the likes of learning, fatigue and boredom.
2. **History.** This is similar to that above but is more concerned with environmental changes that may impact the research, rather than with those changes in the subject. These environmental changes will impact significantly on longitudinal studies and those where the data collection (e.g., interviews or experiments) is conducted over an extended period.
3. **Testing.** The outcomes of a set of tests may be partly attributable to the outcomes of prior tests. The effect of repeated measures may therefore extend over a lengthy period.

4. **Subject mortality.** Subjects may die or, more likely, absent themselves part way through a series of experiments or interviews. This will cause data loss but, more seriously, the absence may be dependent on systematic factors: the least-motivated subjects may drop out of an experiment so that we are left with an unrepresentative group. Brownell (1995, p. 11) observes that in experimental settings, failure to recruit participants, and their failure to re-appear, may be related to the treatment, especially if some treatments are perceived to be less desirable than others!

   Casey (1980) reports different response rates to his survey; those with the greatest information load were least likely to respond. Similarly, in archival studies on company cases, the failure, merger or acquisition of some companies means that they will be removed from the sample and will be under-represented in time series studies.

5. **Experimental mortality.** The passage of time can have numerous effects on the subject companies. For example, the companies may have grown, merged or shifted from public to private status. A survivorship bias can arise because failed companies have been omitted from an all-company analysis due to unavailable information.

6. **Instrumentation.** Identical materials, instructions and procedures must be used throughout the study, other than where treatments are being deliberately varied. If an instrument generates different measures of the same thing under different conditions, then this suggests that contextual problems associated with the instrument, and its administration, have threatened internal validity.

   Accidents or poor planning that cause inconsistencies threaten internal validity and may necessitate the 'scripting' of researchers to ensure there are no unintended distractors, as well as the use of the same researcher for the administration of all experiments or interviews in a cluster to minimise researcher bias. When the impact of several different variables is being tested simultaneously, many different versions of the test instrument must be prepared. These must be checked meticulously (in the same way as examination papers) because a missing page, or a case error, can invalidate the whole experiment. If errors of this type are to be made, then they must occur at the pilot stage with student audiences – the researcher cannot afford the risk of wasting a dataset with a practitioner audience because of an untested instrument. For example, in the pilot stages of Smith and Taffler (1996) it became apparent that while one treatment included 'mean' data, another had included both 'mean' and 'standard deviation' data. Differences in group performance might have been attributable to data differences, rather than to the substantive treatment, which would have invalidated the results of the experiment (even though in practice subjects would tend to ignore standard deviation data). The instrument had to be corrected and the whole experiment re-run to ensure the integrity of the outcomes.

7. **Selection.** Subjects should be allocated randomly to cells (groups for comparison purposes), but this may be either impossible to achieve in practice or insufficient to control for all the variation in the sample. For example, an assumption that control for experience and gender, prior to allocation to random cells, will be sufficient may be unrealistic if there are wide disparities in individual characteristics among the subjects. In some studies, notably in field research, randomisation may be impossible to achieve. In exceptional others it may be undesirable. Thus, Cheng et al. (1998) chose experimental participants based on their cognitive style in order to explore the impact of this particular characteristic. A selection bias can also interact with the variables of interest to the researcher in a manner that precludes the ability to generalise to other subjects. For example, if the selection of subjects was based on volunteers interested in accounting topics, then these individuals may be expecting change and reacting to it, in a way at variance to that of a 'real-world' decision-maker who is less interested in accounting. Where the subject selection is essentially a self-selection process (very commonly the case in accounting experiments), then the personal experiences and expectations of those so attracted may not correspond with those of a more representative set of individuals.
8. **Statistical regression.** There will be a statistical tendency for successive results from individuals to regress towards the mean. This should be reflected in the interpretation of the pattern of successive results over time. Thus, where subjects generate extremely high/low scores in one test, they are less likely to do the same on a subsequent one; they are more likely to regress towards the mean.
9. **Imitation of treatments.** Where subjects can communicate with each other it is possible that we will not achieve independent responses. This can be overcome in experimental settings by asking subjects not to collaborate, or by manipulating instrumentation orders to preclude useful comparisons, or by providing 'filler' processes to occupy potentially disruptive subjects.
10. **Resentful demoralisation.** Different treatments may cause different levels of motivation among subjects, with a consequent impact on outcomes. For example, those subjects receiving feedback on one experiment may be better motivated for subsequent experiments, than those who do not.

Threats to external validity

Christensen (1994) specifies three forms of external validity concern.

1. **Population validity.** Research findings should be generalisable to other people, companies, countries, and/or cultures, as appropriate. But many research samples are not representative – often because of access problems

in field study settings – making extension of the findings problematic. Trotman (1996) suggests that case variability constitutes a significant external validity threat, especially if the number of extreme cases employed in an experiment is at variance with those confronted in practice.

2. **Ecological validity.** Research findings can be generalised to other situations and environmental settings. This requirement spurs the search for greater reality in experiments and increased attention to the context of the experiment. Data limitations often provide a constraint in experimental research since the design is accompanied by a removal of the richness of the data. Analytical requirements also necessitate the use of data categorisation methods in order to use the simplified treatment levels of ordinal variables (e.g., yes/no, high/low, structured/unstructured) in order to accentuate the observed variation in ANOVA-type analyses.
3. **Temporal validity.** Research findings can be generalised across time.

Brownell (1995, p. 13) identifies three further external validity threats.

1. **Treatment/selection interaction.** Managers of organisations usually choose those members of staff who will participate in research; they will usually be the 'better' performers and they will always be those who have some spare time!
2. **Treatment/setting interaction.** Without replication (which would normally be impossible to achieve) we cannot generalise the findings from a particular group of people in an organisation to other people at different levels or in other locations or different organisations.
3. **Treatment/treatment interaction.** Where prior treatments appear to impact on subsequent ones (in within-subject experimental designs), there may be a conditional relationship which requires a redesign of the research model.

Experiments are designed to achieve internal validity in order to confirm or disconfirm existing theories. Such theories may then be generalised to provide some external validity. However, the experimental conditions may be so extreme that they bear little resemblance to actuality, making generalisations difficult. Replications would help in the verification of outcomes, but the accounting literature is extremely loathe to publish replications of this nature, so they tend not to be conducted.

The possibility therefore exists that research will produce sample-specific findings. This can be the result of a non-representative sample selection process or the fact that the unusual combination of various internal validity problems has led to peculiar results for the specific research project.

Quasi-experimental research

Experiments conducted in the field still appear very rarely in the literature, largely because of the constraints imposed by access, ethical considerations and even trade unions. Such constraints were by no means as restrictive in 1924, when a series of experiments commenced which have become one of the most celebrated examples of experiments-in-the field, though largely for the wrong reasons. The Hawthorne Experiments (Mayo, 1933; Roethlisberger and Dickson, 1939) remain a classic indicator of what can go wrong under experimental conditions when planning does not adequately anticipate the range of possible responses. The experiments were conducted in the period 1927–32 in Western Electrics' Hawthorne works in Chicago, and examined a supposed causal relationship between productivity, as the dependent variable, and physical working conditions. The latter was defined in terms of a combination of illumination, temperature, humidity and frequency of rest periods. Productivity levels prior to the experimental period were used to match two groups – an experimental and a control group – and alternative treatments were applied to observe the impact of varying the four dependent variables. But output in the experimental group increased for all treatments (even those meant to induce lower productivity), and the output of the control group increased without the application of any treatments. Further manipulations, including lengthening the working day and eliminating rest periods, also failed to constrain the observed increase in productivity! Uncontrolled mediating variables associated with employee behaviour (notably attitudes, values and norms), together with the conduct of the experiment, apparently colluded to produce outcomes the opposite of those anticipated. A number of important issues emerge from this, which have wider implications for experimental research:

- The 'matching' process employed leads us to believe that we may not be comparing like with like. Randomisation in the assignment of subjects to groups is much preferred, especially if we wish to conduct any statistical analysis.
- The failure to isolate experimental and control groups under identical experimental conditions provides a threat to internal validity.
- Initial experiments included only six female personnel (even though in excess of 40,000 people in total were employed at Hawthorne), suggesting that the samples are scarcely representative.
- Potentially unco-operative subjects were excluded from the experiment so that we may be dealing with a biased and unrepresentative group.

- Experimental mortality is a serious problem with longitudinal studies of this kind; absenteeism and terminations produce discontinuities of data collection.
- A large number of intervening variables associated with employees' attitude to work and each other have been overlooked (e.g., the influence of job security).

The peculiar experimental outcomes may in part be explained by the sort of behaviour referred to by Weber and Cook (1972); the employees are so glad that someone is taking an active interest in their work that they seek to deliver what the researchers are looking for – namely, increased productivity. The 'surprising' outcomes experienced by the researchers, subsequently termed the 'Hawthorne Effect', have become associated with occurrences where the behaviour of groups derives from them seeing themselves as 'special' in novel situations. The outcomes were therefore attributable to the experimental conditions, with mediating variables obscuring the causal relationship under investigation and threatening the internal validity of the experiment. We may argue that the researchers should not have been 'surprised' by these outcomes because Myers (1924, p. 28) had earlier reported almost identical outcomes in a substantially similar context. In an accounting parallel, Smith and Taffler (1995) report unexpected subject behaviour during the conduct of their experiment in that respondents all chose to adopt decision-making heuristics to reduce the decision-making burden under conditions of information overload. They had been expected to re-evaluate decision outcomes separately, and hypothesis testing required them to do so, but none did. The analysis of the experimental findings and conclusions are therefore restricted in a manner which may not have happened had the authors anticipated this occurrence and redesigned the experiment accordingly. Further, they require the additional, and unplanned, analysis of the different timesaving heuristics being employed. Svenson (1979), in the psychological literature, had foreshadowed just this kind of behaviour in information overload situations.

The observed threats to validity associated with a 'Hawthorne Effect' can conveniently be split in three ways:

- indexicality: the variation of day-to-day behaviour by individuals in a longitudinal experiment;
- experimenter effects: bias introduced from the researcher behaving differently from day to day and between groups, providing unintended information which results in different outcomes;
- subject mediation: differences associated with individuals interpreting the task requirements differently.

The implications for these three effects on accounting research may be described respectively as:

- external validity threats because subjects will not behave similarly in different contexts, or even in the same context at subsequent time periods;
- internal validity threats associated with using different researchers for the administration of the experiment, when they are neither scripted nor approach the experiment in identical fashions (see Rosenthal, 1966);
- internal validity threats, derived from assumptions about the personal characteristics, abilities and motivation of research subjects, and their understanding of the experimental task and instrumentation.

The kind of outcomes that were observed in the Hawthorne studies provide a warning of the care that is needed if we are contemplating the search for greater external validity in experimental settings. Quasi-experimental designs, associated with moving out of the laboratory, aim to reduce the artificiality of the situation. A trade-off of internal validity is inevitable, for lack of control over both extraneous variables and participating groups may make manipulation of treatments impossible, which may, as a consequence, threaten internal validity. Quasi-experimental designs usually preclude the random assignment of subjects to treatment groups. This absence necessitates sound theory at the outset, and the use of pre-tests prior to the manipulation of treatments, so as to establish a base for the evaluation of judgements. The use of a control group – with no treatment – is also helpful, but may be impossible to achieve.

Additional problems, over and above those already noted for true experiments, arise in the field because of the need to collaborate closely with the organisation that is providing access. Managers are often not very concerned about research, but are interested more in findings that may be useful to their organisation. They may treat the study as a problem-solving exercise or consultancy assignment, and be indifferent to research niceties. They may insist on their allocation of employees to treatment groups (without random assignment) and view groups that are not testing out some improvement opportunity as a waste of resources (which will mean no control groups either).

Such difficulties may persuade some researchers to adopt a covert approach, leading them to introduce manipulations without management or employees being fully aware of them. Apart from the ethical problems associated with such actions, the outcomes are often not worthwhile because the manipulations to the treatment may have to be so small, in order for them to be introduced covertly, that they go unnoticed and will produce no effects. Worse, if trust is lost with the sponsoring agency, then access may be denied and employees may refuse to co-operate appropriately.

Given the difficulties, it is hard not to read some reports of quasi-experimental studies and form the impression that the research has run away from the researcher! The findings are interesting, but the researcher is struggling to justify both their methods and their degree of control over the whole research study.

Detailed planning at the design stage is imperative in experimental settings. Even so, it is very difficult to be 100 per cent certain of covering everything that may cause a difficulty, meaning that additional experiments may be necessary to correct flaws. Where access to subjects is highly restricted, based on either time or availability, this can be problematical, and at the very least will delay the analysis.

7

Survey Research

CHAPTER CONTENTS

Survey methods are often criticised as being the 'poor man's experiment' because of their inability to assign subjects randomly to treatments, and their consequent inability to rule out rival hypotheses. Brownell (1995, p. 31), among others, recognises the internal validity threats, but he suggests that survey studies can be designed to minimise such threats, while optimising their external validity benefits. He emphasises the need for good theory in order to underpin the specification of causal relationships.

Surveys can be conducted via mail, telephone, e-mail, internet or face-to-face interview. There is still very little literature in the accounting domain with respect to e-mail or internet-based research. The dominant methods remain the mail survey and face-to-face interview, each with its own advantages and disadvantages. These two methods provide the focus of this chapter, with the emphasis on the former.

Mail surveys

Young (1996, p. 55) highlights the decline of mail survey methods in management accounting, after a period in excess of 25 years, during which it has

been the pre-eminent research method in use in the discipline. He attributes the decline to three major factors:

1. A growing interest in alternative forms of research which can provide richer data sources.
2. The increasing difficulty of having mail survey studies published in major refereed journals.
3. Doubts about the usefulness of survey research in accounting since it has failed to yield a cohesive body of knowledge concerning accounting and control practices, despite 25 years of trying to do so.

Brownell (1995, p. 60) also questions the predominance of survey questionnaires in management accounting research, and the reliance on instruments taken from organisational behaviour. Together, these two provide potentially serious weaknesses, so that questions of context are 'typically handled either badly or not at all' in experimental survey studies in accounting. A fourth point may be added to the above list – the difficulty of achieving adequate levels of response to mail surveys, despite following stated guidelines. For example, Brown et al. (2000) report a response rate of only 13% despite using a personalised sample with support from a sponsoring organisation.

Young's analysis of mail survey studies published in major journals (*Accounting Review*, *Accounting Organizations and Society*, *Journal of Management Accounting Research* and *Behavioral Research in Accounting*) over the period 1985–94 identifies a number of common difficulties:

- low target populations (average only 207);
- low numbers of respondents (average only 146);
- few studies using follow-up procedures to increase the sample size;
- absence of analysis of non-response bias;
- absence of studies using both subjective and objective measures of performance;
- absence of the use of sampling procedures; and,
- failure to collect both factual and psychological data within the same study, making it impossible to link practices with behavioural variables.

Such failings lead to the impression that survey research is rarely good research, and is therefore only exceptionally equipped for publication in the most prestigious outlets.

Young (1996, p. 67) identifies seven improvement opportunities:

1. **Research programmes to establish a framework for research**. There remains the pressing need to develop a coherent body of knowledge in management accounting, to match those in finance and financial accounting. Opportunistic research to date has limited developments in this regard so that 'budget impact' remains the major area in the discipline that has been extensively researched.

2. **Sampling methods leading to more powerful theory testing.** Random sampling is usually not practicable so convenience sampling predominates. Such criticisms apply to accounting research generally, and make the application of standard statistical testing fraught with danger.

3. **The use of Dillman-type methods to achieve larger sample sizes.** Follow-up procedures (Dillman (1978) suggests sending two reminders) and sponsorship by participating organisations may increase response, but will incur extra costs and a lack of research control. If we are to guarantee anonymity in the conduct of the survey, as ethical guidelines require, then the survey response should not include the name or position (or perhaps company) of the respondent, to prevent any link between a particular individual and a specific survey response. This will increase the costs associated with follow-up reminders since letters will be sent to both respondents and non-respondents (often much to the annoyance of the former). There is the temptation to adopt unethical numbering systems, colour coding, or even invisible ink (!) to identify participants, while appearing to deliver anonymity, but the use of a dual-response mechanism (anonymous questionnaire *plus* named postcard) usually satisfies both cost and ethical concerns.

4. **Addressing the issue of non-response bias.** The absence of any reference to non-response in many published papers, or the ubiquitous footnote to the effect that non-response was not considered a problem, are of concern. They suggest that no serious attempt has been made to examine the issue.

5. **Moving away from outmoded survey instruments.** It is still common to see papers published in 2002 using standard instruments like those developed by Mahoney et al. (1963) for self-reported performance; Milani (1975) for budgetary participation; and MacDonald (1970) for tolerance of ambiguity. Their age suggests that it must be possible to generate a more relevant current instrument, but the incentives to do so are slim. A new instrument is vulnerable and needs extensive testing, and there are no corroborative studies using the same instrument. We observe the classic trade-off between reliability and construct validity, the choice of a well-accepted and reliable instrument that may only approximately capture the construct of interest. Thus Merchant (in Brownell, 1995, p. 149) uses the LBDQ (leader behaviour description questionnaire) adapted by Aiken and Hage (1968) from Halpin's instrument developed in the 1950s. The LBDQ measures two dimensions of leadership (consideration and task orientation) while Merchant (1985) uses it to measure 'detail orientation', even though he subsequently suggests that this particular instrument may have been less than optimum.

6. **The development of surveys on the basis of improved organisational knowledge.** If the survey instrument does not correspond with the

'language' of the firms involved, then response will be limited because of the perceived irrelevance of the survey.

7 **Moving away from subjective self-reported measures to more objective evaluations.** Young observes an almost total absence of, arguably, more objective superior ratings. However, other authors (notably Brownell, 1995, p. 44) suggest that such criticism of self-rated performance is overstated, and that superior ratings are just as likely to be in error because of the range of subordinates under the control of a single supervisor.

Design and planning issues

A number of fundamental questions need to be answered at the design stage.

1 **What sort of survey are we contemplating?** The requirements of the research question and the impact of cost differentials, for example, will both be important in determining whether a conventional mail survey is appropriate, or if surveys conducted by telephone, e-mail or through the internet would yield superior and/or more cost-effective outcomes. Mail questionnaires allow a large enough sample to reduce sampling error to acceptable levels, at considerably lesser costs than either telephone or face-to-face interviews. In addition, mail surveys provide no opportunity for interviewer bias, a potentially serious problem in both face-to-face and telephone interviews.

2 **What sort of respondent are we targeting?** It will make a great deal of difference at the planning stage, depending on whether we are targeting the population in general or a very specific proportion – for instance particular professional groupings, or even CEOs. The narrower the grouping, the more essential it is that we have up-to-date mailing details of the individuals who are to be contacted. If we wish to contact very specific members of the population (e.g., sets of twins for environmental studies where we wish to eliminate the impact of heredity), we may have to advertise for participants.

3 **What questions do we want answers to?** It may appear obvious, but it helps in this respect if we have carefully specified research question(s) and hypotheses to direct expected responses. Too often in research papers and dissertations it appears that the survey has been conducted first, perhaps because of the opportunistic availability of access, without the research question really having being thought through. This quickly becomes apparent when key questions that should have been asked are

found not to have been asked in the subsequently developed research questions. Roberts (1999) suggests that best practice in the development of instruments and questionnaires dictates that an extensive review of related instruments is undertaken first, and that where instruments need to be purpose-built or adapted, pilot testing is required to address issues of relevance and wording.

4 **What response categories are we contemplating?** For example, are we asking for opinions, judgements or knowledge? Are we setting questions which are closed (requiring yes/no, Likert-scale responses or tick-a-box type answers) or open (allowing a considered narrative response)? We need to address these issues early on or they can come back to haunt us. If we are expecting a narrative response, for instance, then we must provide the respondent with enough room to give it; if we have a mass mail-out, we need an efficient coding system to deal with all the closed questions; if we are asking for knowledge, then questions must refer to items that we can reasonably assume a respondent to know without having to search or look up the details. One of the most serious criticisms of survey research (e.g., Chua, 1996) is that the questions asked are often so complex that the survey questionnaire ceases to be the most appropriate method of data collection.

5 **What sequence of questions should we pursue?** There are varying opinions as to whether the easiest and shortest questions should be at the beginning of the questionnaire or at the end. Some authors (e.g., Parker, 1992) suggest that short and easy questions should be used at the beginning, leading to the meatiest questions in the middle of the survey, followed by relatively shorter and easier questions towards the end in order to encourage completion of the whole survey document; others (e.g., Bryman, 2001, p. 117) suggest that early questions should clearly be relevant to the research topic and should not address personal issues like age, experience and educational background. At this stage we must also consider whether there will be an order effect, that is, would we have generated different answers by ordering the questions differently? If we think that this is likely, we should rerun the survey with a smaller pilot audience to determine whether or not our fears are justified.

6 **What layout of the survey instrument?** Most authors concur that the survey should not be too long, but that, more importantly, it should be made interesting and relevant to the target audience. Long questionnaires are more cost-effective, but only if they are actually returned! The optimum length depends on the format of the survey instrument (e.g., the desire to leave white space, or the requirement to provide gaps for narrative inputs), but should not normally be greater than four pages for the general population. Specialist groups may tolerate something slightly

longer, perhaps of the order of six pages. Maintaining interest and motivation mean that the typical respondent should be able to complete the instrument in less than 20 minutes.

7 **How do we select the sample?** This is important and a weakness in many papers, where the issues seem to have been brushed aside – probably because they have not been adequately addressed in the first place. There are several key considerations: Do we know the size of the population and its constituent items? In many accounting research projects the answer to this question is no. As a result scientific methods of sample selection are precluded, and we need to appeal to the opportunistic or convenience samples so common in the literature – even though they may be 'dressed up' to look like something more systematic. If we have a known population, we will probably have a readily available sampling frame (a stock exchange yearbook, for example, for companies, or an electoral register for individuals). We can then sample randomly from this population (perhaps using a random number generator), or choose every nth item to deliver the required sample size, or stratify the population according to its characteristics in order to ensure that we deliver a representative sample. There are mathematical formulae for calculating the required sample size to deliver the necessary statistical accuracy of estimates, but it is usually easier to return to the research question and hypotheses. We should be able to specify the tests we intend to perform, and the number of ways that the data will be split; we can identify all the cells of the analysis and we would like at least ten (20 is better) items in each to give us confidence in conducting the intended statistical tests. If we cannot adequately resolve order issues at the pilot stage of the survey, then the requirement for multiple versions of the final instrument will expand the required sample size by the same multiple. For example, we do not want to be in the position where we want to test for gender effects, say, with data on individuals, and then find we have too few females in the sample (which happens more often than it should with accounting data). Similarly, when testing for industry effects with company data, we may have too few retail representatives, though with some countries (e.g., Australia and New Zealand) the populations themselves may not be big enough to yield the samples required to conduct tests of all desirable relationships.

Pilot testing

Extensive piloting of the survey instrument is essential to demonstrate that it is capable of generating the required responses from the target audience. Although surveys are often tested on academic colleagues or undergraduate students, it helps if members of the target population, who have been excluded

from the sample, are used to gauge its viability. The pilot instrument should be an advanced draft of a document which adequately represents the progression of the research from abstract concepts, through the development of valid constructs, to the identification of reliable individual questions. A number of important issues will arise at this stage, which must be addressed satisfactorily:

- the questions must be clear, simple and easily understood;
- the questions and covering letter should be targeted towards the respondent's viewpoint so that they are clearly relevant to the target audience. Any jargon or industrial terminology employed should be technology-specific so as to improve response rates;
- the choice of words must be careful, avoiding slang, abbreviations and any terms with potentially ambiguous meanings;
- there should be no 'double-barrelled' questions because if more than one answer appears to be sought, confusion will result;
- double negatives should be avoided, as they will frequently be misunderstood;
- however, wording reversals should be employed to prevent respondents from unthinkingly 'ticking the right-hand box', say, without paying due reference to the precise meaning of the question;
- respondents must have the knowledge and skills to equip them to answer the questions. This is currently an issue in auditing research which is targeted at subordinates in accounting firms – the levels of complexity in some surveys place unrealistic demands on those responding to the survey;
- those questions which are incapable of producing reliable answers must be eliminated. Thus, questioning individuals in social research about their sexual behaviour, gambling habits, drug or alcohol abuse are unlikely to produce accurate responses. In accounting research, questions relating to fraudulent practices, dysfunctional behaviour, income, and even position, may elicit misleading answers. Similarly, questions relating to religion may cause difficulties in cross-cultural research;
- attention to the time taken by pilot respondents in completing the survey should give an early indication of whether individual questions, or whole sections, need to be pruned.

To improve the reliability and validity of individual questions the entire questionnaire should be evaluated at the pilot stage prior to conducting the survey proper. A large pilot study would also allow an evaluation of the reliability and validity of the measures to be used.

Data collection

A number of further considerations arise when we focus in more detail on the collection of the data.

1. **A relevant and up-to-date mailing list is essential.** Use of an existing mailing list may require sponsorship by a host organisation (e.g., one of the professional accounting bodies). The 'cost' of sponsorship may be recognition of the host organisation in any publication and/or some loss of control in the conduct of the survey in that the host organisation handles completions and returns so that there are no guarantees as to exactly who has completed the survey! Alternatively, the purchase of a reputable database may be required. This can be expensive for a narrowly focused mailing list. The development of a mailing list of one's own from 'scratch' is extremely time consuming and labour intensive. It is also an ongoing involvement because vigilance must be exercised in the maintenance of the mailing list. There is nothing likely to cause more consternation among survey recipients than if one of the named targets is already deceased.

2. **The survey should target specific named respondents.** There is a wealth of evidence which suggests that surveys that are addressed to the 'occupant' or the 'manager' or some other unnamed individual are those most likely to be consigned to the waste bin. The research literature (e.g., Dillman, 1978) suggests that surveys should be targeted by both name and position, and that if there are any doubts in these respects they should be confirmed in advance by telephone (e.g., through the company exchange operator) before delivery of the survey. Dillman further suggests the use of a clear covering letter, ideally on headed notepaper and signed by a recognised dignitary, which should provide unambiguous instructions, a guarantee of confidentiality, and a demonstration of the importance of the survey and its relevance to the respondent. Merchant (1985) customises his research instruments by varying the technological terms to fit the target audience. In so doing he ensures the relevance of the survey to the recipient, increasing the response rate to 95%, but jeopardises reliability through variable instruments.

3. **How do we record the answers?** This should be established early to make the most of the media employed. If we are dealing with a mail survey, then manual methods will predominate. However, if we have verbal (i.e., interview or telephone response) or written responses (i.e., narrative answers in a manual survey, or e-mail or internet responses), then opportunities exist to conduct a detailed qualitative analysis of the narrative through the content analysis of the text, even though this may have to be transcribed from tape recordings.

4. **Feedback to respondents?** The offer of aggregated results to respondents may provide an incentive which encourages completion of the survey. This is often more successful than the offer of prizes or nominal rewards for return of the survey. In any event, a letter of thanks to respondents is good manners, and may elicit an increased willingness for further

involvement. For example, it may encourage respondents to make themselves available for follow-up interviews, to provide both clarification and detail. If response is still the major objective, then stamped addressed envelopes, preferably with real stamps rather than bar codes, are preferable. Dillman (1978) recommends the sending of two follow-up reminders to elicit further responses, as well as a careful monitoring of holiday or busy-business periods so that these may be avoided for the survey distribution (e.g., avoidance of company surveys close to the financial year-end, the end of the tax year, Christmas or Easter).

5 **Organisation.** It is better to think ahead. At the planning stage we should be aware of the coding necessary for closed answers, and also of the methods of analysis to be employed. Ideally, the responses should readily be transferable to spreadsheets for manipulation, or to SPSS input, to facilitate more detailed analysis.

6 **Non-response problems.** The biggest concern in survey research is lack of response. If respondents are unrepresentative and response rates are extremely low, then doubts will arise about the validity of the findings and the potential for biases being introduced. Response rates of less than 25% are common in accounting research; the question that is difficult to answer is whether respondents differ significantly from non-respondents. Non-response is only a problem if we can demonstrate that there are systematic differences between respondents and non-respondents, and that such differences will impact on the findings. This latter condition may be difficult to demonstrate. Merchant (1985) uses a 'postcard' method both to guarantee the anonymity of respondents and to distinguish between respondents and non-respondents. Participants are asked to complete the questionnaire and a separate postcard, which are independently mailed back to the researcher. Assuming that respondents do indeed return both items, then the identities of non-respondents will be known and their characteristics can be compared with those of the known respondents. In the absence of such a device we are forced to estimate the characteristics of non-respondents by proxying their characteristics from those respondents who were the last to respond after the final reminder. The implication is that these last-minute, almost reluctant, respondents will resemble those that did not bother to respond at all.

Dillman's (1978) Total Design Method (TDM) pays particular attention to the reasons for non-response:

- a wrong address in the first place, or incorrect postage attached, resulting in non-delivery;
- the unopened letter discarded because it looks too much like a circular or other junk mail;
- the delivery is to an inappropriate person, who fails to forward it appropriately;

- there is no motivation to complete the survey when opened, so it is discarded;
- the recipient cannot understand the completion instructions and/or the survey content;
- the survey instrument is temporarily 'shelved' because of time pressures;
- the return-address has been misplaced, so that even if completed, the survey instrument does not get returned.

He tackles each of these problems by ensuring that the instrument is 'right', that it reaches the correct person, and that incentives for completion (prizes or monetary rewards) are provided. He also institutes a detailed series of follow-ups:

- one week after the initial survey – with a reminder postcard (some researchers suggest that this first reminder should be sent as soon as three days after the original mailing!);
- three weeks after the initial mailing – a new covering letter and questionnaire;
- seven weeks after the initial mailing – a third covering letter and questionnaire.

The whole survey process is considerably lengthier and more expensive than those we may normally contemplate as a result, but Dillman can point to his response rates as a rejoinder.

Measurement error

Andrews (1984) specifies three kinds of measurement error: bias, random errors and correlated (or systematic) errors. He suggests that better questionnaire design will help to overcome the most serious effects of these errors, including:

- the use of as many answer scale categories as possible, consistent with parsimony and survey length;
- a 'don't know' option where appropriate;
- keeping the number of items grouped together (termed the 'battery length') small;
- the use of comparative scale questions where possible to provide an explicit judgement base;
- the use of a linear rating scale, with only the extreme categories labelled;
- care in the length of both 'battery' introductions and questions (recommended as between 16 and 24 words for the former, and more than 16 words for the latter).

- positioning easy and less important questions at the beginning and end of the questionnaire.

Measures of reliability

If we are to draw valid conclusions about the relationships under investigation it is important that the variable measures are both reliable and valid. Multi-item scales are preferable to single-item scales for two major reasons:

- many constructs represent complex concepts, and this complexity is better addressed by asking more than one question;
- a number of related items can increase validity because of the possibility that individual questions may be misinterpreted or misunderstood, or they may contain missed wording reversals.

There are several commonly used methods of measuring reliability:

- Test–Retest Reliability Coefficient: the same instrument is completed twice by a single group of individuals within a short time period. A high correlation between the responses suggests reliability, but the outcome the second time around may have been influenced by the first completion.
- Split-Half Reliability Coefficient: the instrument is completed by one group and the correlation is measured between two halves of the instrument (usually between first and last to examine the effects of fatigue). But many errors are attributable to reversals in wording not picked up by respondents, suggesting a preference for the random selection of the two halves for testing. If the correlation between the two halves is high, then this suggests high reliability on an internal consistency basis; but the longer the test, the higher the reliability, and the way the split is performed is likely to impact on the measure of consistency/reliability. The logical extension to this method is then the generation of an Average Inter-Item correlation coefficient ($\bar{r}$) based on the average of all possible pairwise items.
- Cronbach's Alpha Coefficient is the most widely used measure, especially for newly developed instruments, and overcomes the splitting problem, though its value is still wholly dependent on the number of items (n) in the instrument. Thus sensitivity to the number of items should be evaluated to demonstrate reliability.

$$\alpha = \frac{n * (\bar{r})}{(1 + (n-1) * \bar{r})}$$

where $\bar{r}$ would be calculated from averaging $\frac{(n!)}{(n-2)!\,2!}$ correlation coefficients.

Thus for a four-item construct, there would be six correlation coefficients to average to generate ($\bar{r}$). An alpha of 0.8 is normally deemed to be satisfactory, though figures slightly lower than this may be acceptable. Since the Cronbach alpha depends on the number of items included, the more items the higher the Cronbach coefficient. Very high coefficients, resulting from too many similar

questions, may therefore reflect redundancy in the instrument. A sensitivity analysis of the alpha to deletion of successive items will reveal whether we have a parsimonious set which is consistent with reliability.

Interview methods

Many of the problems associated with self-completion mail questionnaires are also applicable to interview methods, but there are also additional and inevitable 'people' problems because of the interaction between interviewer and interviewee. However, if we can alleviate the difficulties, interviews offer greater opportunities for dealing with more complex and wide-ranging issues than do conventional mail surveys. A number of interview formats are common in the accounting literature, and are addressed here. Case studies and field research are considered separately in the following chapter.

1. **The structured interview.** This is the format which most closely resembles that of the self-completion mail questionnaire. Opportunities for interviewer bias are restricted by seeking a common context: the same questions, in the same order, with the same cues and prompts permitted, and all within a specific, closed-question framework. The use of closed questions makes the coding of answers easier and has advantages for the subsequent analysis. Closed questions also eliminate the opportunities for error associated with open questions, as well as the chance of 'missed' questions where order differences are permitted. But closed questions also sacrifice the comparative advantage of the interview method by failing to include the flexibility and richness of response offered by open-ended questions. In the accounting literature, Lowe and Shaw (1968) and Onsi (1973) provide examples of the adoption of a structured interview approach.

2. **The semi-structured interview.** This format allows a series of questions to be asked, but in no fixed order. Additional questions may also be asked, as the interviewer sees fit, to examine associated issues that arise in the course of the interview. Lillis (1999) provides an example of the use of the semi-structured approach.

3. **The unstructured interview.** This format commences with a series of topics for discussion, rather than specific questions to be asked. It may develop into a directed conversation, with the interviewer able to adopt a 'free-wheeling' approach, as long as the required topics are all covered. The actual words and phrases used may therefore vary significantly between interviews, but this approach may put interviewees at their ease sufficiently to induce them to make disclosures that would not have

emerged under different conditions. The unstructured interview approach is illustrated by Merchant (1985).

Several areas of concern may arise, which may cause the outcomes of the interview process to be questioned:

- Poorly worded questions may cause confusion or misunderstandings among interviewees. It is possible that interviewer and interviewee have different interpretations of the terms used or the emphases accorded them – an extension of the sort of 'measurement of meaning' issues raised in the accounting literature by Houghton (1987, 1988) and Hronsky and Houghton (2001).
- Memory problems among interviewees can make instant responses unreliable. This may be associated with questions which are far too demanding for this mode of investigation.
- Questions may be asked in an inconsistent manner or a particular interviewer may behave differently over time and between respondents. Significant differences may exist between interviewers despite levels of training, guidelines and standardisation of questions. This within- and between-interviewer inconsistency may result in biased outcomes if, say, overly sympathetic or aggressive attitudes are linked to a particular group of interviewees.
- Problems may arise in the recording and processing of responses. This occurs particularly with open-ended questions, where interviewers may misinterpret or embellish responses in the course of hurriedly transcribing what has been said.
- Non-response bias can arise, as it does with mail surveys, but refusals are compounded by absenteeism. Face-to-face requirements may necessitate frequent call-backs (analogous to the reminder letter) and their success may depend on the gender, dress-code and non-threatening attitude of the interviewer.

A special category of problems for interviewers arises with regard to concern for ethical issues. While the covering letter demonstrates the purposes of the research in mail surveys, and the return of the instrument itself constitutes implicit 'permission to participate', the interview situation is different. Some ethical codes will even insist that the interviewee 'signs-off' at the completion of the interview to verify the content of the responses. Particular points that should be addressed include:

- interviewers must clearly identify themselves, their status, who they represent, and the purposes of the interview;
- interviewees should be aware of how and why they have been chosen to participate;
- the confidentiality and anonymity of responses must be emphasised;
- voluntary participation is paramount so that interviewees are free to withdraw their co-operation at any time;

- there should also be the opportunity for interviewees to ask questions of the interviewer. Interviewers need to be careful here because if they are off their guard they may be induced to reveal the research questions under examination. Such disclosures may threaten the validity of responses from subsequent interviewees.

Well-designed surveys allow the relations between the variables of interest to be rigorously studied. However, internal validity is not as strong in surveys as it is in experimental research, though the extra realism provides higher external validity. Since surveys often embrace many cases, which necessarily limits the depth, they are often subject to criticism relative to case studies.

In this chapter we hope to have specified both the advantages of surveys and their limitations. Highly structured questionnaires, restricted to non-complex questions, cannot probe issues in depth, and there are no opportunities either to respond to queries or ask appropriate follow-up questions. These limitations are not significant as long as the questions are well formulated and relatively narrow in scope; they can then provide valid responses to research questions. As with all of the research methods we examine in this volume, the question of theory cannot be understated. Good theory underpins everything that we do, and allows the development of clearly defined constructs. The demonstrable development of reliable research instruments to collect data is one of the great strengths of survey research in accounting.

There is much confusion in the literature between surveys, field studies and case studies. For example, Merchant (1985) described his research as a 'field study' and Smith (1994a, 1994b) described his as 'semi-structured interviews' when they may more aptly have been described as 'unstructured interview' and 'field study' respectively. Brownell (1995, p. 156) suggests that a distinction between 'surveys' and 'field studies' may be determined by the degree of structure in the questioning, which is a necessary, but perhaps not sufficient, condition for the distinction. These issues are explored in the following chapter.

8

Fieldwork

CHAPTER CONTENTS

Over the past 15 years, accounting researchers have been urged to shift their study of the role and function of accounting to its natural contexts. This has been particularly the case in the management accounting literature, led by the seminal works of Hopwood (1983) and Kaplan (1983). While the call for fieldwork has been persistent, the penetration of this research into top-ranking journals, particularly in the USA, has been very limited (Foster and Young, 1997; Shields, 1997). Foster and Young (1997) argue that few studies meet the criteria of high-quality field research routinely applied in other disciplines. Similarly, Shields (1997, p. 10) attributes the lack of publication of case and field study research to several factors, with 'lack of knowledge about how to do good studies' leading the list. Kirk and Miller (1986), among others, emphasise the consideration of the validity and reliability of field studies, in particular:

- whether it presents a consistent interpretation of the data;
- whether the empirical observations justify the extent of the conceptual generalisations;
- whether rival explanations for the observations have been considered and/or eliminated; and
- the replicability of the findings, that is, whether different researchers in different time periods are likely to generate similar findings.

Internal validity is a problem because there are no opportunities to assign subjects randomly to treatment groups in the field, as we would under

experimental conditions. This makes theory even more important to the process, because it forms the major means of eliminating competing explanations for different observations. Brownell (1995, p. 77) calls for attention to be given to the development of more systematic procedures for the assessment of internal and external validity and reliability in field studies. The nature of qualitative research, and the absence of structured survey questions, means that the dataset will not include Likert-type multiple-items measures. Because traditional forms of reliability assessment are not possible, the demonstration of coherent and valid constructs presents a significant challenge for the researcher.

Chua (1996, p. 220) notes that typically, 'good' anthropological fieldwork requires 'a long stay in an alien place'. Such an approach is rarely adopted by accounting field researchers. Most field studies are cross-sectional rather than longitudinal (Ferreira and Merchant, 1992) and the latter is exceptional because extended in-depth access to commercial organisations is usually very difficult. Given that most studies will be associated with doctoral study, the availability of grant and/or scholarship support for a limited period means that on-site data collection for an extended period will be impracticable. Access of any kind will usually be opportunistic (e.g., see Buchanan et al., 1988) and may be subsequently denied (Young and Selto, 1993). Preston (1986), Chua and Degeling (1993) and Chua (1995) provide examples of longitudinal ethnographic studies in single organisations that facilitate the exploration of process issues; Merchant and Manzoni (1989) provide an example of a cross-sectional, multi-case design that may be potentially useful for theory testing.

Three distinct forms of field research can be distinguished, namely, complete participant, complete observer and participant–observer.

1. **Complete participant.** This type of research can take one of two forms:
   - participant as observer: the observer's role as a researcher is concealed from the participant organisation (e.g., Rosenhahn, 1982, as a member of staff in a mental hospital);
   - observer as participant: where the observer is an existing member of the organisation and conducts research in that organisation (e.g., Ezzamel and Bourn, 1990).

2. **Complete observer**. The observer has no contact with the subjects being researched. Such an anthropological approach is almost unknown in the accounting literature. At the extreme (particularly when associated with ethnographic studies), involvement in the research process will be as a non-participating observer. However, unless it is covert, even observation at this level is likely to impact on the behaviour of individuals, and hence on research outcomes.

   So active involvement, rather than detached observation and measurement is the norm in action research projects. Indeed, such involvement is

facilitated by a researcher who has the practical expertise and theoretical knowledge to bring a new perspective to organisational issues.

The nature of the exercise requires a high level of trust and collaboration between researchers and management (non-researchers); the latter may resist any imposition of a structured research design, preferring action research that will result in outcomes of clear benefit to the organisation. For the internal participants, a feasible solution to a taxing problem will be the appropriate measure of success. They may not be interested in testable hypotheses and causal methods so that reporting at two levels may be required: an internal report of a consulting nature, which will focus on problem solving, and a research report which forms part of a dissertation, or which ultimately yields a refereed paper.

The ethnographic extreme has its origins in social anthropology, with the observation of tribes, and in natural science, with the observation of the behaviour of animal troops. A shared knowledge is used to account for patterns of behaviour and personal relations, with potential implications for humans in business settings. Emphasis is on observation, but may be complemented with semi-structured interviews, though not active participation. A series of questions arises as to the manner in which ethnographic studies may be conducted. Consider them, for example, in terms of our wish to observe processes of an executive nature, perhaps at Board level:

- Participant or non-participant observation?
- Overt or covert observation?
- Ethically defensible deception under covert conditions?
- Direct or indirect observation?
- The reliability of secondary data?

Of course, where sensitive issues are under discussion, we may have to settle for secondary data, or a sanitised version of events. Such material, for example, minutes, may well provide no detail of discussions or the contributions of individual participants, power structures or group dynamics.

Covert observation has the advantage of allowing manipulations (e.g., changes in behaviour or role) to be introduced, without their being attributed to the research process. It may also be impossible to conduct the research were the researcher to be so identified. However, Ditton (1977), reported in Bryman (2001, p. 93), amply illustrates the problems associated with covert operations. He was exploring fraudulent practices in a bakery, and was regularly forced to excuse himself to make notes on Bronco toilet paper. The time spent on toilet breaks eventually induced him to 'come out' regarding his covert research! Coffey (1999) provides an example of overt participation in his study within a UK accounting firm.

3. **Participant observer.** The most common scenario, this is where the researchers interact with the members of the organisation in a collaborative venture. It is often termed 'action research'. The participation is usually active on both sides, and is more than either a consultancy project or an in-company problem-solving exercise. The research is guided by theory in examining the change process, allowing the anticipation of consequences and outcomes. The process should have external validity in that it should produce generalisations which have relevance to other organisations. However, we should be reminded here of the problems alluded to in Chapter 1 during our discussion of Heisenberg's Uncertainty Principle, in that the intrusion of the researcher, as participant observer, may make it impossible to conduct measurement in an unbiased manner.

   The research thus constitutes a planned intervention in the normal processes of an organisation, to observe the consequences, and measure and monitor how these differ from those expected from theory. Researcher intervention (either as observer or active participant) is an essential ingredient of the research design, as is the generation of results which will both be of interest to the host organisation and make a contribution to knowledge in the area.

   Although it may be possible to identify groups and structure, their membership (as we would do in an experimental setting) attempts to assign individuals (e.g., by matching or randomisation) would normally be avoided on the grounds that it would destroy the nature of the research setting.

Case study methods

There is frequently much confusion over the terminology of fieldwork compared to case study research. The term 'case study' usually implies research confined to a single unit of analysis, which might be a single department, company, industry or even country. The scope of the case could still, therefore, be broad, but the 'single unit' focus means that it is much narrower than might be embraced by the term 'fieldwork', where the latter would encompass more general studies of social activity in the field. Ryan et al. (1992, p. 114) distinguish five categories of accounting case study:

1. **Descriptive** – where current practice is described in terms of the procedures adopted. The studies may seek to confer 'best practice' or 'successful'

labels to particular sites or companies (e.g., Peters and Waterman, 1982; Smith, 1994c).

2. **Illustrative** – where the researchers explore the implementation and outcomes associated with innovative practices (e.g., Dikolli and Smith, 1996; Kaplan, 1984; Kaplan and Norton, 1992).
3. **Experimental** – where the research concerns the conduct of an experiment in the field, whereby new treatments are applied to sub-units of the site. Examples of this type of research remain very rare in the accounting literature; the most famous and notorious example in related literatures remains the Hawthorne studies (Mayo, 1933).
4. **Exploratory** – where the researchers conduct a preliminary investigation about how and why particular practices are adopted. Such research may be difficult to publish in itself, unless it makes a clear contribution to theory or method.
5. **Explanatory** – where research seeks to provide convincing explanations which justify practice choices and facilitate the development of theory. However, opportunities or attempts to generalise the findings are rare.

Yin (1984, p. 39) emphasises that case study sites must not be chosen because they are *representative* in some way, because researchers should not be concerned with producing statistical generalisations. Rather, he suggests that theoretical generalisations are more important. He compares case-study research with experimental research, in that with the latter we start with a theory, devise an instrument to test the theory, and then choose a representative sample of cases, which are to be subject to alternative experimental treatments. In case research, the case becomes the instrument through which we test the theory, thus the focus of the case must be on whether it can be used to test the extent to which existing theory provides good explanations. A 'representative' case may not be able to do this. Similarly, he suggests that it is impossible to assert how many cases should be included in a multi-case research design until the nature and scope of the cases has been determined.

Case studies would normally collect data from multiple sources, including some or all of the following: documentary evidence, interview data, direct observation and participant observation. Merchant (1985) and Simons (1990) provide examples of studies which have incorporated data from multiple sources. The adoption of multiple-methods is termed 'triangulation' and offers the opportunity to access different sources both for a common research method (within-method triangulation) and with different methods (between-methods triangulation):

- within-method triangulation, which combines different researchers, different interviewers and different survey sites; and
- between-methods triangulation, which combines different results from, say, interviews, survey and archival data collection, and may include both quantitative and qualitative approaches.

It may be possible to offer alternative views of the same phenomenon through a process of 'triangulation', which may increase the validity and reliability of the research. Thus Lillis (1999) uses a semi-structured interview schedule in conjunction with a structured questionnaire to derive the benefits of quantitative and qualitative methods. But she notes that the semi-structured interview method that she uses is inevitably subject to the intrusive effects of interviewer bias, both during the interview and in the analysis of transcripts. Bias is potentially introduced in the coding and interpretive phases. The researcher ultimately decides how each sentence in the transcript will be coded and, more importantly, interprets the 'meaning' associated with selected sections of text in terms of the theoretical constructs. Consistent and valid coding and interpretation of transcript data are absolutely key to the reliability of this analysis. A means of reducing bias is to use multiple researchers in both the coding and interpretive phases.

Chua (1996, p. 227) emphasises that field research is not 'mere story telling', but rather an attempt to build a theory which leans heavily on existing theories and literature. However, Merchant (cited in Brownell, 1995, p. 150) suggests that in the context of field research that explores relationships between controls and management performance, 'there is no such thing as dependent and independent variables in the real world'.

Field study researchers do not have the equivalent of the Cronbach alpha, from survey research, nor do they have the control and treatment groups of laboratory experiments to demonstrate their attention to reliability and validity. Field research, by design, will never have the statistical basis for establishing construct validity that is common with other research methods. Critics and reviewers ask why focused field research should have construct validity requirements that are any less demanding than research addressing the same issues with different methods (e.g., a survey questionnaire). Arguably the concerns should be the same, but the issues are rarely so because fieldwork provides the opportunity to collect both richer and more complex data. The critical reader of research of all kinds should look for clear meanings of the constructs that describe the concepts under study.

The qualitative analysis protocol

Where the researcher and the researched interact there will be opportunities for bias. Silverman (1989) identifies two extreme forms of field–researcher bias: an unwillingness to collect and report quantitative data, and a tendency to convert all qualitative data into quantitative data. Mason (1994) emphasises that quantitative and qualitative data are complementary sources that should be combined to take advantage of the richness of the qualitative findings, and the potential rigour and increased credibility of the quantitative findings.

In addition to the potential interviewer-induced bias in the *collection* of qualitative data, the *analysis* of qualitative data is potentially subject to significant bias since it relies on interpretations and classifications imposed by the researcher. Qualitative data is also vulnerable because of the absence of established techniques for ensuring that data analysis is both complete and impartial. NUDIST (Non-numerical Unstructured Data: Indexing Searching and Theorizing) is helpful in this regard. It is a qualitative analysis package for coding raw interview transcripts, and it associates the sentences in the transcript with one or more pre-defined themes. Completeness should not therefore be a problem, but impartiality still is because the researcher will have already established the categories for search.

The use of a systematic analytical protocol such as that developed by Miles and Huberman (1994) enhances confidence in the impartiality of qualitative analysis because:

- it provides a chain of evidence from transcripts to the results of analysis;
- it ensures that all cases are used in the evaluation of data propositions, preventing interviewer-based elimination that may introduce unintended bias;
- it provides an analytical framework within which hypotheses can be tested.

The field researcher attempts to use a small set of case data to illustrate and support more general, theoretical arguments. However, irrespective of the research method used, the ability to make broad generalisations from a single study is necessarily limited. When evaluating the external validity of a field study, the results should be considered in the context of other studies that examine similar questions in different settings.

Jönsson and MacIntosh (1997) argue that ethnographic studies have been marginalised in the accounting literature by the focus on agency-based 'rational' accounting theory studies in the USA, and 'critical' accounting theory studies in the UK. They express a preference for extensive field studies in actual companies, rather than approaches which are theory-laden but virtually free from empirics. Chua (1988) and Puxty (1993) provide extensive overviews of the then state of ethnographic research in accounting, and Jönsson and MacIntosh (1997) echo their view that ethnographic studies must be much more than 'good story telling'; they recognise that neutral studies are impossible and that appropriate regard to politics and theory should be shown.

Silverman (1985) suggests a typology comprising three alternative approaches to ethnographic research:

- **Cognitive anthropology:** research which focuses on the individual's competence in communicating within the culture under study (i.e., the way in which the actors are able to communicate and behave, thus making them acceptable to the rest of the group). Dent's (1991) study of the shift of

British Rail's culture from an engineering one to a finance/accounting one provides an example of this approach in the accounting literature.

- **Symbolic interactionism:** research concerning the manner in which actors change the culture of organisations with which they are involved. Preston's (1986) study of management information processing in a large plastics division, as a participant-observer, provides an example of this approach.
- **Ethnomethodology:** research which is less concerned with the communication and interaction approaches (above) and more concerned with the social behaviour of the actors involved and their interpretation of outcomes. Jönsson's (1982) study of budgetary behaviour provides an example of this approach in the accounting literature.

Critical research, on the other hand (see Chapter 8 for more examples), focuses on power and class structures in interpreting meaning. Accounting systems are seen as a means of both achieving control and exploiting the workforce. Jönsson and MacIntosh (1997, p. 376) question the 'critical' approach in that its conclusions are always known in advance – 'the cart always comes before the horse'. The outcomes are always attributable to an exploitative capitalist system (see Ezzamel and Willmott, 1992). For example, Tinker's (1980, p. 147) study of financial accounting statements in Sierra Leone, which was able to conclude that 'accounting served to reinforce the institutionalised subordination of black wage labour over the entire period', had not necessitated a site visit or an ethnographic study.

We can enter into the field situation with very little theory to cling on to; the intention may be to develop theory as a result of the initial case findings. A number of alternative methods have been devised to accomplish this feat, including grounded theory (Glaser and Strauss, 1967), analytic induction (Denzin, 1970) and case study research (Bloor, 1978). Without attempting to argue over the superiority of alternative labels, the following common characteristics emerge:

- a definition of the phenomenon under investigation with initial expectations;
- a hypothetical explanation of the phenomenon;
- a sequential examination of cases to determine the extent to which they fit the hypothetical explanation;
- where variation is observed, a modification of the definition of the phenomenon, the case characteristics, and reformulation of any hypotheses in the light of observations;
- an examination of a small number of incidences of the phenomenon (cases) to identify shared features and points of variance;
- speculation on the reason(s) for the continued observation of 'deviant' cases;
- a continuation with the reformulation exercise for successive cases until a universal model can be established which fits a large number of observations.

Certain features of the participant–observer process can be identified that may be thought to constitute disadvantages, and to endanger the validity of the research outcomes and subsequent theoretical developments:

- closeness to unique events;
- limited opportunities to classify and generalise the data;
- the unrepresentativeness of the sample;
- the actual presence of the researcher, which may of itself distort observations;
- the treatment of unusual observations as typical;
- subjective personal observations which may have limited validity;
- a host site which may block access to research subjects.

A number of common characteristics are apparent:

- the researcher's norms, values, code of ethics, assumptions;
- the researcher's impact on problem specification, research strategy and methodology;
- the political context within which the researcher operates;
- resource constraints which hamper the researcher;
- contingencies/opportunities during the course of the research, particularly that of access to funding and host organisations.

Grounded theory has become an increasingly popular method with accounting case researchers, though its acceptance by journal editors has been somewhat slower. Its growing importance justifies separate treatment of its characteristics and advantages (see below).

Grounded theory

Glaser and Strauss (1967, p. 3) argue that 'theory that inductively develops out of systematic empirical research is more likely to fit the data and thus is more likely to be useful, plausible and accessible'. Pure grounded theory analytical approaches are designed to manage and control the potential bias in building theory from empirical data, providing a method which is primarily non-quantitative and designed to find the latent or embedded meanings in data. In a pure grounded theory analysis, theory emerges during the analysis of data; the emergent theory is tested constantly against further theoretically sampled empirical data.

Grounded theory has been increasingly adopted as the preferred qualitative approach in accounting field study environments. However, the subsequent divergence of views between Glaser and Strauss has caused confusion over the meaning of the terms involved and their implications for acceptable procedures. As a result, alternative forms of 'grounded theory' have been developed:

- The basic Glaser and Strauss (1967) approach, embellished by Glaser (e.g., Glaser, 1992), which emphasises an individual approach and personal style;
- The Strauss and Corbin (1990) approach, following Strauss (1987), which is much more structured and prescriptive – and arguably more acceptable to positivist researchers.

A number of other variants have also emerged (see Kools et al., 1996), most notably in the nursing discipline. The rigidity of the Strauss and Corbin (1990) approach was unacceptable to Glaser and other traditionalists because they appeared to be forcing the data into models which did not permit alternative interpretations. All forms of grounded theory have been questioned by positivists on the basis of a lack of external validity, a factor which has contributed to a restriction in the number of journal editors to whom grounded theory approaches are acceptable.

Laughlin (1995) emphasises the importance of the 'researchers' in grounded theory studies in that they are fundamental to the discovery process and the validity of the research, which is dependent upon a shared meaning of events by both researcher and researched. Parker and Roffey (1997) stress the crucial role of language in grounded theory research as a significant discriminator relative to positivist studies. They too point to the crucial interplay between researcher, the researched and data in generating quality research with unique advantages.

Gurd (2002) observes the problems associated with 'labelling' in the reporting of grounded theory research, noting that papers that adopt grounded theory approaches, and report as such at the thesis, working paper and conference paper stages, frequently omit specific 'grounded theory' references when they are finally published. He notes the publication of Soin (1995) as an example. There are other examples, particularly in the management accounting literature, that appear from their content to be adopting grounded theory approaches, but do not admit as such explicitly. The interpretation must be that the confusion concerning what is 'really' grounded theory affords the opportunity for discrimination among journal referees and editors. Clearly it is safer for authors not to specify the adoption of a particular grounded theory approach for fear of generating opposition among referees. Doctoral supervisors must be particularly careful in their choice of external examiners to ensure that a common understanding of grounded theory terms exists between themselves and their candidates.

Verbal protocol analysis

Slovic (1969) provides the foundations for the use of verbal protocols as a method of data collection in judgement exercises. Subjects perform exercises

in which they think aloud, detailing exactly what they are doing, in terms of information being used and the manner of its use. A tape recorder would normally be used during this exercise. The content of these verbal protocols provides the basis for the mapping of the decision process and problem-solving behaviour for each subject. The resultant flowchart will usually be complex, with numerous feedback loops, and will allow the development of conditional algorithms for decision-making. The accuracy of the resultant models can often be measured by comparing the model predictions with the actual subject outcomes, in a 'man vs. model of man' way. Verbal protocol analysis (VPA) has rarely been used in the accounting environment. It provides the opportunity for generating very rich datasets, but the resource involvement is great: only one subject at a time can be evaluated, and the researcher must monitor the conduct of the exercise and prompt the subject, where necessary, to ensure that he or she maintains a constant narrative throughout the decision process.

Larcker and Lessig (1983) use retrospective protocols of decisions, after the completion of an experimental task, to improve on regression models of investment decision-making, originally based on decision accuracy. Bouwman (1984) gives one of the earliest examples of the use of contemporaneous VPA in the accounting literature, when he compares the decision processes of expert and novice decision-makers. Anderson (1985) and Harte and Koele (1995) provide evidence which suggests that contemporaneous verbal protocol analysis can give reliable information about pre-decision behaviour. Anderson (1988) presents a verbal protocol analysis of the examination of initial public offerings (IPOs) by financial analysts, with the focus on the search process, decision time, decision conclusions and the information items addressed.

Boys and Rutherford (1984) offer a fascinating example of the use of VPA, examining the use that investment analysts made of current cost accounting information in evaluating firm performance, when provided with both historic cost and current cost numbers. We may anticipate that this approach would confer benefits on investigations of the use of International Accounting Standard (IAS) information. Most recently, Anderson and Potter (1998) suggest that regression analysis and verbal protocol analysis are complementary methods, as each contributes to model development in the analysis of decision-making behaviour. They develop a model of individual decision behaviour based on the outcomes of a verbal protocol analysis, but it should be noted that their very small sample size – which is a common problem in VPA – of only four individuals makes it difficult to generalise their outcomes.

In conclusion, where resources are available to employ more subjects, and to evaluate their deliberations in appropriate detail, VPA provides a powerful alternative means of exploring decision processes, and one which remains vastly underused.

9

Archival Research

CHAPTER CONTENTS

We take a deliberately broad approach to the term 'archival' in this chapter, by embracing the sources used to generate research based on historical documents, texts, journal articles, corporate annual reports, company disclosures, etc. The associated research approaches may therefore range from the fundamental analysis of accounting numbers, through to the content analysis of narratives and critical approaches to the development of accounting theory.

Searching for information can be a time-consuming and expensive exercise so it is important that researchers quickly develop the skills necessary to locate and use sources effectively. Such sources can usually be classified as *primary* (e.g., original research results published for the first time) or, more usually, *secondary* (e.g., information that has been aggregated and/or reworked in databases). However, such a classification may not be clear-cut because a company annual report may be deemed a primary or secondary source depending on the identity of the user! Such sources can usually be accessed directly, or relevant references sought, through keyword and author searches of library catalogues, abstracts or internet databases. As we suggested in Chapter 1, a critical attitude should be adopted to the research process, and this should apply just as much to accessing data sources as to subsequent stages. We need to be able to evaluate the suitability to purpose of the dataset: Is it up to date? Is it from a reputable and authoritative source? Has it been gathered using reliable methods? Can we access the material in a timely and economic manner, given the constraints of our research budget? The latter is becoming increasingly problematical for extensive online databases,

for without generous educational discounts many such sources would remain inaccessible to academic researchers.

Foster (1986) identifies a number of problems associated with data collection from secondary sources, in both cross-section and time-series studies.

Cross-section data

1. Data may exclude some current companies. This may be a particular problem if multiple databases are being used which do not overlap completely, so that some companies fall 'between the cracks'. In any case, small companies may not be included if there are size 'hurdles' specified for their inclusion. The same principles would apply to those companies which are not actively traded on stock markets. These conditions may also lead to the exclusion of private or foreign-owned companies. A common reason for such exclusions is the non-availability of the data. Particularly annoying in this respect is the absence of data for subsidiary companies where there is no requirement for them to report separately from the parent.
2. Data may exclude non-surviving firms. Merged, acquired and bankrupt firms will normally be omitted from current databases, necessitating searches from other sources if these are the subject of the research. Much past research in the failure prediction area has been criticised for suffering from a survivorship bias because, by definition, failed companies tend to be omitted from the analysis due to unavailable information.
3. Data may not be right up to date in that the most recent data may not have been incorporated. This is becoming less of an issue with more online and web-based databases operating either in a real-time mode or being capable of uploading information on a daily basis.
4. Data may be incomplete in that it omits some financial items. For example, earnings forecasts, or 'notes to the accounts', may not be there, necessitating the use of alternative sources.
5. There may be inconsistent classification of some financial items across firms. If the database comprises other than camera-copies of original documents, then some assumptions are inevitable in order to produce systematic cross-company classifications. For example, where firms are permitted differences in reporting line items, there will be different levels of aggregation, which may only be separable with arbitrary decisions. Thus, one firm might include overhead expenses in 'costs of goods sold', while another might include overheads in expenses attributable to 'marketing, administrative and general'. Unreliable entries may thus result for items such as 'overhead' where disaggregation assumptions have to be made. These kinds of problem are exacerbated by non-synchronous reporting periods (resulting in large differences both within and between

countries) and the non-uniformity of accounting methods, especially across industries, which makes comparisons difficult because different choices may still be consistent with accounting standard compliance.

6. There may be recording errors, necessitating checks against other comparable databases where feasible, and necessitating the use of simple internal validity checks. For example, computing the mean and standard deviation of items allows all of those outside the range of two standard deviations either side of the mean to be identified and questioned. Similarly, simple comparisons of quick assets with current assets may reveal basic errors. Industry classification poses a particular problem here because there is no single, accepted definition of 'industry' and different databases may adopt alternative classifications. Although 'product group' or 'production process' would normally form the basis of classification, without reference to some external regulatory classification problems may occur.
7. The nature of disclosure is expanding all the time, making it more and more difficult for researchers to be confident that they have captured the most reliable and comprehensive sources. In the financial reporting environment most studies still rely on the content of the corporate report, but increasingly newspaper sources are being used because they provide a more timely medium. Reuters Business Briefing (RBB) is probably the most detailed source of company news items available in UK, though it is not widely used for academic purposes. The Financial Times Index (UK) and Wall Street Journal Index (USA) provide popular alternatives (see also http://www.bloomberg.com/). Brookfield and Morris (1992) use the McCarthy Information fiches (now available on CD-ROM). Internet and e-mail disclosures represent additional, relatively untapped sources, potentially important because there is a wealth of evidence that companies are disclosing information through these means to investment analysts prior to its availability to the stock market.

Time-series data

1. Structural changes may have taken place in the company or the industry, making comparisons between time periods fraught with danger. Internally, these may be due to mergers, acquisitions or divestments; externally, they may be attributable to new government policy, deregulation, new products, new competitors or technological change.
2. Accounting method changes, particularly those associated with voluntary choices or switches, may make the financial numbers from successive periods difficult to reconcile. Where this constitutes deliberate obfuscation, it is a particular cause for concern.

3. Accounting classification issues may occasion different corporate interpretations being placed on particular items, perhaps again to cloud the communication issue. Thus a firm may elect to consolidate the results of a subsidiary in one year, but not the next, even though there appears to have been no material change in circumstances between periods. Similarly, the flexibility in reporting the timing and amounts associated with accounting for 'extraordinary items' and 'goodwill write-downs' frequently necessitates adjustments being made in data if a comparative base is to be maintained.

Even if the research project being conducted would not normally be termed 'archival', the points above have implications for the use of any documentary materials to be used to support the other research methods addressed in earlier chapters:

1. Where the database is in the form of a mailing list to support survey research, failure to update it regularly will mean that the list both excludes some target persons and includes some who are either dead or have moved away. Such errors and omissions can cause both bias and irritation.
2. Where the database is a journal listing that forms the basis of our literature review, we have a number of potential problems. The journal may not be available online at all, and will be excluded from all databases; this still applies to many accounting journals which are published by individual universities rather than through professional publishers. Even where they are available, online selected journals may only appear in specific databases – we may need to access multiple databases to track down the required references. 'Old' papers are still not available in an electronic form through most databases, although the databases are becoming more comprehensive in their coverage, with deep back runs. If we need to access seminal works from before the late 1980s, then we may have to resort to a hardcopy print format (see the literature search discussion in Chapter 3). Similarly, the most recent of papers may not be immediately available either; there is nothing quite so frustrating as having access to a title, and perhaps even the abstract, of a must-read paper only to realise that the whole paper will not be available for another two months. Beware, too, of the existence of the notion of a 'whole' paper because sometimes the online version will omit all the figures and references (fortunately this is becoming less of a problem with the use of Adobe Acrobat and the predominance of PDF files).
3. We have to beware of making unwarranted inferences from archival sources, especially where there is the danger that we may not be comparing like with like. Context differences may explain many of the apparent contradictions and inconsistencies in the findings of comparative pieces, making it imperative that we return to the original sources

wherever possible. Indeed, Brownell (1995, p. 140) attributes many of the problems of accounting research to the fragmentation that means comparisons are difficult to make with confidence: namely, different studies using different methods and instruments in different locations.

The validity trade-off in archival research

An archival study will normally have more external validity than experimental or simulation approaches because of its reference to empirical data. But dangers will arise if our selection process (e.g., for company data) is flawed, so that it results in the generation of an unrepresentative sample. This situation will be exacerbated if we employ 'matching' procedures in the research design (typically matching on size and industry) because there will be no guarantees that the findings are not industry-specific, or that they may even be case-specific to the group of companies selected.

Libby (1981) suggests that econometric studies using archival data are essentially experimental in nature. They may be used to answer similar questions to those addressed by experimental studies, even though the opportunities for variable manipulation are limited. While laboratory experiments often manipulate treatments and infer causality, many archival studies search for association and systematic movement between variables of interest. Although an association, rather than causation, is being observed, internal validity concerns still exist. For example, Wallace (1991) specifies the internal validity problems associated with financial statement research, particularly those concerned with 'instrumentation' and 'history' – concerns which will also be relevant in other financial accounting fields.

With respect to instrumentation, Wallace suggests that there are always questions of what exactly constitutes an 'accounting change'. Technical details become critical in the instrumentation process. If different information sources are used or even different personnel to collect data from annual reports, measurement differences may arise which threaten the validity of outcomes. Similar problems of instrumentation arise in failure prediction research, since a variety of definitions of 'bankruptcy' have been used in past research. As Wallace observes, not only are there different types of bankruptcy, but there are questions as to how reorganisations, restructuring of debt, and technical non-compliance with loan covenants are to be treated. If different definitions are being used in the source data or by fellow researchers then internal validity threats will arise. Houghton and Smith (1991) provide an excellent example of why researchers should be wary of comparing the findings of different studies, if they are not prepared to check the detailed definitions employed. The definition of failure in their study included 'subject to stock exchange investigation' – a very wide definition, which is unlikely to coincide with that used in most other associated studies.

With respect to history effects, changes in bankruptcy law, reporting requirements and accounting policy over the period of interest would all affect the comparative findings from archival searches of company data. The absence of adequate controls for the impact of such changes is a cause for concern. The response of researchers is often to use a matched sample that tries to control for extraneous factors. But which factors do we match on? Another problem with this approach is that the selection process precludes any assessment of the importance of, say, size, industry or capital structure, where we have chosen to match on these factors. In addition, measurement issues mean that we are not sure we have matched correctly. For example, do we match size on assets or number of employees? If we select assets, just how close does the match have to be to be ruled acceptable – $1k, $10k, $100k, $1m, $10m? Such measurement issues may prove material.

Content analysis

Content analysis is defined as a method that uses a set of procedures to make valid inferences from texts. The inferences are about the sender(s) of messages, the message itself, or the audience of the message, and the rules for the inferential process vary with the interests of the investigator (Weber, 1985). Statistical studies of literary style, particularly those that solve disputes about the authenticity of writings, date back to the 1850s, and are well illustrated by Mosteller and Wallace's (1963) model, which allows a distinction between disputed authors of the Federalist Papers based on the incidence of the words *whilst*, *upon* and *enough*. Similar approaches were adopted by Osgood and Walker (1959) and Stone and Hunt (1963) to distinguish between fake and genuine suicide notes, based on their reference to things and persons. The analysis of word patterns and sequences to detect hidden messages came to the fore during wartime, with their use in exposing propaganda in political speeches (Berelson, 1952; Laswell, 1948).

Content analysis has traditionally been applied to the analysis of archival data, but is becoming increasingly popular in the analysis of interview transcripts. Typically, quantitative methods have been applied to archival data and qualitative methods to interview transcripts. Where quantitative methods have been employed, they have usually been limited to the manifest characteristics of text (e.g., the number of occurrences of words, or the number of words relating to particular themes). The quantitative results in the form of variables referring to particular words and themes are then available for statistical analysis.

More recently, the techniques have been applied to the qualitative analysis of open-ended survey responses with the aim of corroborating survey data. In these applications, content analyses may examine latent characteristics of the data such as the underlying meaning of the phrases used (Holsti,

1969). A further issue relates to the connection between the manifest and latent content of a narrative. Content analysis rests on the belief that it is possible to go behind the text as presented and infer valid hidden or underlying meanings of interest to the investigator (Weber, 1990, pp. 72–6). Content analytic procedures that restrict themselves to manifest content alone would thus be of very limited value. Salancik and Meindl (1984, p. 243, footnote 2), however, argue that whether or not the attributions expressed are the 'true' beliefs of the authors is irrelevant.

Two alternative generic approaches to content analysis are usually taken where quantitative analysis is contemplated: 'form orientated' (objective) analysis, which involves routine counting of words or concrete references; and 'meaning orientated' (subjective) analysis, which focuses on analysis of the underlying themes in the texts under investigation.

In the managerial literature, Bettman and Weitz (1983), Staw et al. (1983), Salancik and Meindl (1984), Clapham and Schwenk (1991) and Abrahamson and Park (1994) all adopt a content analysis approach to explore the causal attributions made by firm managements in their letters to shareholders to explain or account for company performance. Jones and Shoemaker (1994) provide a general overview of empirical accounting narrative analytic studies. Kelly-Newton (1980) adopts a content analysis procedure to the measurement of themes in her analysis of the general comments section of a sample of replacement cost footnotes examining management reaction to disclosure requirements. Ingram and Frazier (1980) conduct a content analysis of firm environmental disclosures, and also report an explanatory study linking narrative disclosures with firm performance across three industries (Ingram and Frazier, 1983).

Abrahamson and Amir (1996) study the information content of the president's letter to shareholders, and highlight the importance of textual portions of annual reports to investors. Bowman (1984) uses the number of occurrences of the word 'new' in the president's letter as a measure of managerial risk in addressing questions of firm strategic risk and uncertainty. He also emphasises the advantages of content analysis as an unobtrusive measurement since the statements are written for purposes and audiences different from those constituted by content analysts. There is little danger that the measurement process will confound the data (Weber, 1990, p. 10).

D'Aveni and MacMillan (1990) use content analysis of shareholder letters to analyse the differential strategic responses to demand crises by the top managements of surviving and bankrupt firms. Successful firm managements are distinguished by their focus on critical success factors in their output environment (e.g., customer needs and demand growth), whereas failing firm managements deny crises, look inwards and focus on the short term.

Building on Frazier et al. (1984), McConnell et al. (1986), Swales (1988) and Yoon and Swales (1991) all use a content analysis approach to explore whether qualitative data found in the firm's annual report can forecast stock price performance. In contrast to Bowman (1984), who focuses on positive

references, Abrahamson and Amir (1996) restrict consideration to negative references only. Tennyson et al. (1990) explore the relationship between the firm's narrative disclosures and bankruptcy using a content analysis approach, but they do not differentiate between positive/negative or good/bad references in their statistical models.

Weber (1990, p. 37) argues that word categories inferred from covariation among high-frequency words are more reliable than themes. However, Krippendorff (1980, p. 63) suggests that for many content analyses, thematic units requiring user judgement in the determination of the hidden messages conveyed in the narratives may be preferable despite application difficulties. The term 'word' is taken to indicate semantically equivalent textual units, including word synonyms, idioms and phrases (Weber, 1990, p. 22), and 'theme' is taken to mean clusters of words with different meanings or connotations that, taken together, refer to some theme or issue (Weber, 1990, p. 37). Smith and Taffler (2000) adopt these basic definitions in their conduct of both form-oriented (word-based) and meaning-oriented (theme-based) analyses. The qualitative content of the narrative is transformed into quantitative variables for subsequent analysis with simple formulae. Thus for words:

$$\text{Word variable} = \frac{\text{Number of common occurrences}}{\text{Total number of words in the narrative}}$$

Ratio variables can then be computed from the narrative for each theme on the basis of the importance of those themes in the narrative. Thus, if a sentence comprises four themes, each is accorded a theme-score of 0.25. The overall score summed across all sentences accorded any particular theme is taken to be indicative of its importance within the narrative.

$$\text{Theme variable} = \frac{\text{Sum of theme scores}}{\text{Total number of sentences in the statement}}$$

Reliability and limitations

Krippendorff (1980, pp. 130–54) warns against the potential unreliability of self-applied investigator-developed recording instructions, emphasising three aspects of the process:

- stability – inter-temporal coding differences in the same coder should be insignificant;
- reproducibility – coding rules should be such as to allow different coders in different locations to agree substantially on their assignments;
- accuracy – the performance of coders should largely comply with a known 'right' answer, although this is frequently impossible to assess in practice.

There is no generally agreed level of performance inter-correlations that are deemed to be satisfactory, but Krippendorff (1980, pp. 146–7) suggests that inter-coder reliability correlations in excess of 80% should be sought. Nonetheless, however careful the researchers are, as Weber (1990, p. 62) emphasises, content analysis is partly an art and depends on the judgement and interpretation of the investigator. 'Texts do not speak for themselves. … The investigator must do the speaking and the language of that speech is the language of theory' (Weber, 1990, p. 80). Researcher bias cannot be avoided.

Another major limitation of content analysis is that it assumes that frequency of occurrence directly reflects the degree of emphasis accorded to words or themes, but this may not always be so (e.g., see Weber, 1990, pp. 71–3). In addition, words or sentences classified in the same category for data reduction purposes may not reflect that category to the same extent (Weber, 1990, p. 72).

Critical analysis

In Chapter 1 we distinguished between positivist, interpretive and critical approaches to accounting research. Much of the content of our earlier chapters has been devoted to 'positivist' perspectives, so it is opportune here to return to the other two approaches, especially since much of the associated literature is documentary, historical and archival in nature, and frequently without an empirical base.

Baker and Bettner (1997) note that this lack of empiricism contributes to the absence of critical and interpretive studies from the top US accounting journals. Jönsson and MacIntosh (1997) question the appropriateness of the 'critical-Marxist' stance adopted by such as Tinker (1980) and Ezzamel and Willmott (1992), especially where their conclusions are largely determined in advance, and without the necessity for fieldwork of any kind. We return to these issues in Chapter 8. At the other extreme, 'interpretive' research conducted in the hospital sector (e.g., Chua and Degeling, 1993; Preston et al., 1992, 1997) is highly empirical and dependent on extensive fieldwork assignments.

Power et al. (2002) criticise the view that accounting represents economic reality. They follow Hopwood (1987) and Hines (1988) in suggesting that accounting is implicated in the creation of that reality. Arrington and Puxty (1991) suggest the need for 'less accounting', but more 'accountability'.

It is possible to identify the development of a number of narrower fields of critical/interpretive research, adopting different perspectives on the distorting nature of accounting, notably:

- **Critical-Marxist:** for example, Tinker (1980); Tinker et al. (1991).
- **Critical-Radical:** for example, Armstrong (1987); Cooper and Sherer (1984); Tinker and Niemark (1987), who adopt neo-Marxist perspectives to suggest that accounting distorts reality by representing the interests of capital.
- **Critical-feminist:** for example, Hammond and Oakes (1992) and Hines (1992), who provide critiques founded in a gendered conception of accounting logic.
- **Interpretive:** for example, Arrington and Francis (1989); Lehman (1999).
- **Critical-interpretive:** for example, Laughlin (1987); Power and Laughlin (1996).
- **Feminist-interpretive:** for example, Broadbent (1998); Gallhofer (1998).
- **Radical:** for example, Sikka (2001); Sikka and Willmott (1997).
- **Radical-feminist:** for example, Hammond (1997).

Such labelling is always fraught with danger and likely to raise the ire of the authors concerned. What it does do is demonstrate the breadth of research being conducted in the critical paradigm, despite the restricted nature of the publication opportunities. Indeed, almost all of the works referenced in this section have been published in just three journals: *Accounting Organizations and Society*, *Critical Perspectives on Accounting* and *Accounting, Auditing and Accountability Journal*.

10

Supervision and Examination Processes

CHAPTER CONTENTS

Research supervision arrangements and examination processes differ quite significantly between countries, but whatever the environment it is important that the candidate has reasonable expectations as to the role of the supervisor(s), the level of guidance to be provided, and the nature of the examination process. Thus in the USA, a supervisory panel will be the norm and, following the completion of an extensive programme of structured coursework, the dissertation component will usually be shorter than those expected elsewhere. An oral examination of the dissertation will often be conducted in front of a large, potentially public, gathering.

In both the UK and Australia, a team of two supervisors is usually appointed, though one of these supervisors may adopt the 'principal' role. The second supervisor may be junior in both age and experience, and may be perceived by the candidate to be 'learning the ropes' in some way. In such circumstances it is desirable that the principal supervisor seeks to involve the second supervisor in all meetings and deliberations, to avoid his or her isolation, and to avoid the development of a one-on-one relationship with the principal supervisor, often at the instigation of the candidate. There may be little formal coursework other than studies of research methodology, and few, if any, assessment hurdles prior to completion of the dissertation/thesis. An oral defence of the thesis (*viva voce*) is the norm in the UK, a potentially fraught exchange in front of a small audience – two examiners (only one of whom is usually external to the university) and a non-participating supervisor. Travel costs to and from the *viva* examination frequently constrain the choice of examiner. This situation is magnified in countries like Australia,

where the distances involved are so prodigious and the pool of potential examiners so shallow that overseas examiners regularly need to be sought. Consequently, it is more normal for there to be no *viva*, though it remains an option in exceptional circumstances. We anticipate an increasing use of video-conference technology in the future, to facilitate more *viva* examinations. As a trade-off, this scenario usually demands the employment of two (sometimes three) external examiners who report independently on the hard-copy document. This arrangement has the advantage of making it possible to choose the best examiners possible worldwide, without the constraint of personal attendance. The downside is that there is no opportunity for oral evaluation of the candidate. Whatever the arrangements, it is vital that candidates know what is expected of them, and how their supervisor(s) can help them to satisfy the examiners at the first attempt.

The role of the supervisor

The assignment of supervisor(s) to candidate may take place in a variety of ways. In the traditional model, a candidate will enrol with a specific supervisor attached, and with (at least) an outline research proposal in hand. The supervisor will usually have been given some say on the suitability of both candidate and project, and be prepared to modify the latter in order to satisfy the mutual objectives of supervisor/candidate. Alternative arrangements include the acceptance of candidates on to an extensive structured programme, without either a supervisor being attached or a detailed research idea having already been developed. This second form has the advantage of facilitating the assignment of the most suitable supervisor to the candidate, but the disadvantage that the candidate may develop a proposal for which there is no suitable supervision available! However, a reality check during the structured programme should help to ensure that candidates do not pursue directions likely to pose insuperable supervision problems.

Where a candidate enters into a programme with specific grant or bursary funding already attached, then a specific project may already have been determined for which supervision is virtually self-selecting, being those academics who initiated the original project proposal.

Lack of supervisory capacity for doctoral study is a serious constraint for most accounting/business departments. There is usually no shortage of doctoral candidates requiring supervision, but incentives for supervisors to supervise are conspicuously absent. This is especially so for experienced supervisors, with a sustained record of on-time completions. Excess demand for the services of these individuals (a demand which continues to increase rapidly because of the growing popularity of DBA programmes) means that candidates have to be realistic in their expectations. For example, it is

reasonable to anticipate that a supervisor will wish to supervise topics within their existing expertise, without the necessity of exploring new and complex literatures outside their normal sphere of activity. It is also reasonable to expect that they will wish to have the opportunity of publishing with the candidate, either during the period of candidature or after completion. A successful supervisor will usually have multiple candidates at different stages of the research process so that two or more candidates complete each year. But as well as excess demand there may be supply constraints in the form of arbitrary 'caps' implemented by the university to restrict the number of candidates attached to any one supervisor. These arrangements make it vital for candidates to sell themselves and their topics effectively, and to be prepared to be flexible in the structuring of their proposals, if they wish to be associated with the most successful supervisors.

Having said that, the 'most successful' supervisor may not be the 'best' choice in every case. The supervisory relationship is a very personal one, and one which will involve a great deal of detailed communication and close interpersonal relations. It helps if supervisor and candidate can get on with each other, and ideally develop a friendly working partnership which goes beyond a master–servant relationship. But all candidates and supervisors are different, making it difficult to dictate the behaviour of both parties. Unfortunately, in a minority of cases, this relationship can break down, often because of the intransigence of either or both. In my experience, problems are most likely to arise when the candidature is well advanced (often with less than one year to go before projected completion). By this time candidates should be confident; they should know their literature better than their supervisor – possibly better than anybody else in the world! At this time they may wish to pursue a direction tangential to the current one, and at variance with that agreed with the supervisor. The latter may be fearful that a new direction is risky and may lengthen the candidature outside a normal completion time. They may be unwilling to loosen the reins and back the new proposals, to the extent that working relationships become impossible. There are arguments on both sides of the divide, and the candidate's wish to innovate is part of his or her own personal development. However, he or she must be prepared to put a time limit on the new venture, so that he or she is prepared to come back into line if that is necessary to ensure timely completion. Where such compromises prove impossible, a change of supervisory arrangements will be necessary, but the candidate must recognise that the choice of alternatives may be limited. What is important is that there is a supervisor in place – even though he or she may not necessarily be a discipline expert – to guide the candidate through the final stages of structuring and writing the thesis prior to submission.

There are a number of distinct stages in the supervision process where the supervisor and candidate have particular responsibilities. There will inevitably be differences on the nature of each's responsibility, so Table 10.1, inspired by Moses (1985), provides a 12-stage check-list in the form of a 5-point Likert scale as a basis for discussion between the potential protagonists so that some broad agreement can be reached.

TABLE 10.1

Expectations of supervisor and candidate

1	It is the supervisor's responsibility to select a research topic	1❍ 2❍ 3❍ 4❍ 5❍	The candidate is responsible for selecting her/his own topic
2	It is the supervisor who decides which theoretical framework or methodology is most appropriate	1❍ 2❍ 3❍ 4❍ 5❍	The candidate should decide which theoretical framework or methodology he/she wishes to use
3	The supervisor should develop an appropriate programme and timetable of research and study for the candidate	1❍ 2❍ 3❍ 4❍ 5❍	The supervisor should leave the development of the programme of study to the candidate
4	The supervisor is responsible for ensuring that the candidate is introduced to the appropriate services and facilities of the school/faculty/university	1❍ 2❍ 3❍ 4❍ 5❍	It is the candidate's responsibility to ensure that he/she is appropriately located and accommodated, and can access all relevant services and facilities for research
5	Supervisors should only accept candidates when they have specific knowledge of the chosen topic	1❍ 2❍ 3❍ 4❍ 5❍	Supervisors should feel free to accept candidates, even if they do not have specific knowledge of their topics
6	A warm, supportive relationship between supervisor and candidate is important for successful candidature	1❍ 2❍ 3❍ 4❍ 5❍	A personal, supportive relationship is inadvisable because it may obstruct objectivity for both candidate and supervisor during candidature
7	The supervisor should insist on regular meetings with the candidate	1❍ 2❍ 3❍ 4❍ 5❍	The candidate should decide when he/she wants to meet with the supervisor

(Continued)

TABLE 10.1

(Continued)

8	The supervisor should check regularly that the candidate is working consistently and on task	1❍ 2❍ 3❍ 4❍ 5❍	The candidate should work independently and not have to account for how and where time is spent
9	The supervisor is responsible for providing emotional support and encouragement to the candidate	1❍ 2❍ 3❍ 4❍ 5❍	Personal counselling and support are not the responsibility of the supervisor; the candidate should look elsewhere
10	The supervisor should insist on seeing all the drafts of work to ensure that the candidate is on the right track	1❍ 2❍ 3❍ 4❍ 5❍	The candidate should submit drafts of work only when he/she wants constructive criticism from the supervisor
11	The supervisor should assist in the writing of the thesis if necessary	1❍ 2❍ 3❍ 4❍ 5❍	The writing of the thesis should only ever be the candidate's work
12	The supervisor is responsible for decisions regarding the standard of the thesis	1❍ 2❍ 3❍ 4❍ 5❍	The candidate is responsible for decisions concerning the standard of the thesis

Although we might reasonably anticipate that supervisors should select the first option in Table 10.1, many supervisors will themselves disagree on what their overall responsibilities are, so that it is important that any supervisor/candidate pairing is clear from the outset of their mutual expectations.

The supervisor should be able to plan and manage the research process so that it produces interesting and innovative outcomes in a timely fashion. Although the candidate may generate the initial research idea, the supervisor must ensure that this topic will produce an 'acceptable' thesis in that it is

neither too big (and impossible to complete on time without overwork or burn-out) nor so small that the examiners will not perceive it to make a sufficient contribution to knowledge. This can be a difficult compromise so that many supervisors will over-compensate slightly to avoid the latter situation. It can cause conflict when candidates are anxious to complete and can point to much shorter PhD theses (especially US ones) for comparison. Supervisors must plan the whole research process to provide a structure and establish priorities to facilitate the development of a clear proposal within six months of candidature; such a process establishes expectations, but also imposes some discipline on the participants, which may also be the source of friction.

It is usual for the candidate to be enthusiastic about his or her project at the outset; lack of enthusiasm at this stage is a cause for concern which may suggest that the student will not stay the course and complete. Phillips and Pugh (1994), especially in their fourth chapter 'How not to get a PhD', provide an interesting and amusing guide as to what a candidate should not do if he or she seeks a timely doctoral completion! Supervisors must monitor levels of enthusiasm to ensure that motivation is preserved and progress is still being made. This can be a serious problem, particularly among part-time candidates, where the competing demands of 'work' mean that research progress becomes extremely slow or even non-existent. The specification of particular targets or milestones during the research process should help. Some candidates respond very well to deadlines. If these are associated with the submission of conference papers relating to work-in-progress, they can be particularly beneficial.

From the supervisor's perspective, the most problematical aspects are those to do with how much of themselves is in the thesis – the sort of considerations associated with points 1, 2, 7, 11 and 12 in Table 10.1 in particular – and these will often be closely associated with the differing skills of the candidate. Thus:

1. It helps if the candidate arrives with a well-developed research proposal because his or her motivation and commitment at the outset is usually greater. But candidates may have very sketchy ideas or be unable to prioritise issues in the literature satisfactorily at the beginning of their candidature. This will mean that supervisor input is great. Many candidates expect the supervisor to more or less specify the project; many supervisors will only countenance candidates in very narrow areas of research so that their projects are effectively supervisor-specified.

2. Some supervisors with narrow perspectives will work with a single methodology, so that theory and approach are effectively pre-selected. Others, myself included, see the development of theory and method as an essential part of the literature review process; an examination of prior relevant work is used to identify alternative theoretical justifications and different possible research methods.

7 The requirement for regular meetings between supervisor and candidate is eminently sensible, though it is not clear what constitutes 'regular' in this context. Some research regimes with well-developed quality assurance procedures will have strict reporting requirements for the regularity, duration and content of such meetings. However, the rapid improvements in information technology, and the equally rapid growth in candidates studying remotely and online, may make face-to-face meetings largely superfluous, especially for able and well-motivated candidates.

11 The intellectual property for the final dissertation manuscript rests with the candidates – the words chosen should be their own. However, some candidates, especially those for whom English is a second language, require a great deal of assistance in the editing of drafts, sometimes to such an extent that they resemble re-writes. It may be necessary to employ a professional editor to help the candidate prepare the final manuscript prior to submission, to ensure that it satisfies the minimum requirements of grammatical accuracy. If candidates are publishing and presenting their findings during their candidature, which is usually desirable, then the supervisor may well be the co-author on a number of papers which embrace the content of the dissertation. It would be unrealistic if the final version of the dissertation did not incorporate the majority of the edits and revisions necessary to generate publishable papers. In this case the final dissertation is the joint responsibility of candidate and supervisor.

12 The supervisor will normally be required to 'sign off' candidates, that is declaring that the work of their candidates is both their own and deemed worthy of submission for examination purposes. It may be possible for the candidate to insist on submission without such an acknowledgement (for example, where the working relationship has broken down irretrievably), but such action is inadvisable, being a significant source of subsequent failure at the examination stage of candidature. To make such a declaration the supervisor must have sighted the final drafts of the thesis, made appropriate recommendations for change, and then sighted the final version to be submitted. Unfortunately, reference to some of the dissertations where I have been involved in the administration of the examination process suggests that this is not always the case. The supervisor should also be actively involved in the selection of external examiners (dealt with in more detail in the following section), by securing the collaboration of suitably qualified individuals. They may illicit the advice of candidates in this process, since their literature searches and conference attendance should have facilitated the identification of eminently qualified examiners. Some supervisors will want their candidates to know the identity of examiners well before the examination, not so that they can contact them (which is strictly disallowed and

would disqualify the examiner), but so that they can ensure that they have appropriately cited relevant works published by the examiner.

The importance of the role of the supervisor in the research process cannot be underestimated. Excellent performance can bring great credit to both university and candidate. Inadequate performance can cause a plethora of difficulties, which can result in slow progress, inappropriate choice of examiners, subsequent failure and potential legal ramifications. Research administrators have the vital task of providing supervisors with the freedom that exploits their gifts, while instituting an unobtrusive monitoring mechanism that protects the candidate's learning experience.

Examiner profiles

The supervisor will start the process of sourcing a suitable external examiner about three months prior to the expected submission date of the final thesis. Such a time lag is essential because it may be difficult to identify individuals who are both willing and capable of examining. This can be particularly problematic in areas like finance and banking, and for dissertations which span more than one discipline.

The external examiner will be an active researcher, currently supervising his or her own doctoral candidates, and with a healthy track record in the discipline area of the research project. This is a fundamental requirement, and will usually be the minimum permissible qualification for the examiner to be nominated to act in the first place. He or she will have published recently in areas within the scope of the dissertation, and this publication record will suggest that he or she is sympathetic to the research perspective adopted by the candidate. This is an important requirement, with which the supervisor should be well versed. Errors in this regard can produce enormous problems further down the track. For example, it may not be sufficient to identify a 'management accountant' as an examiner; the dissertation may be a qualitative, interpretivist piece while the selected examiner is an avowed positivist! Such a match would not be in the best interests of the candidate.

Given the expertise of the examiner in the area of research, it is highly likely that he or she will have published research close to the dissertation topic. It would be remiss, even rude, of the candidate not to have cited relevant work published by a potential examiner. In just the same way as we would not like to annoy unnecessarily a reviewer of a journal paper, then we should try not to irk an examiner through being unreasonably critical of his or her prior work. We must not forget that even examiners are typically sensitive academics, and they would prefer to see their work described as 'pioneering contributions' or 'seminal literature' in the area, even if the findings have

subsequently been superseded. Ideally, examiner, supervisor and candidate will all have corresponding views of what constitutes good research.

In securing the collaboration of the external examiner, the supervisor must follow some of the sound principles of 'impression management' identified above in securing access to research sites. Flattery is often a key component of the initial approach adopted by the supervisor, since there is otherwise little motivation for the examiner to be involved. The financial inducements are usually negligible and, other than for inexperienced examiners, there is little professional kudos in being involved in the examination process – just a lot of work! It is important, therefore, that the dissertation topic is in an area of interest to the examiner, where the examiner has demonstrated his or her competence, and where he or she might be deemed something of an expert. A well-written abstract will help to 'sell' the dissertation topic to a potential examiner by demonstrating its importance and the contribution it will make. This is an important opportunity for candidates to sell themselves and their work to an examiner before the formal process commences. Good first impressions at this stage can be helpful. Unfortunately, most candidates write poor abstracts. They spend insufficient time on them and tend to try to cobble together pieces from elsewhere in the dissertation rather than write something original. Examiners will see through this and will not be impressed. Where an examiner has been identified as the best possible, it is worth making the effort to secure their co-operation.

The examination process

The formal evaluation process commences when the dissertation manuscript is delivered to the examiner. Impressions created in the first five minutes of perusal can be important to the examiner's ultimate opinion so it is important to take some care in this respect.

- Ideally, the dissertation should be a single volume, even where university regulations permit the submission of multiple volumes. Dissertations perceived as being too long give the impression that the supervisor may have exercised insufficient control over the progress of the research.
- The volume should be immaculately presented. It should be professionally bound in hardcovers (even where regulations permit submission in cheaper, flimsy covers). The professional look does not cost much, and is well worth the impression it creates. Well-spaced gold lettering on both the cover and the spine add to this impression.
- The volume should be free of grammatical and spelling errors. Accurate editing and proof-reading should ensure that this is so, making it doubly unfortunate that the front and contents pages of many theses contain glaring errors – even to the extent of spelling errors in the title! Tables, in

particular, should be rigorously checked for both spelling and numerical errors, especially since routine proof-checking will normally entail 'reading around' the tables so that these are frequently overlooked in the checking process.
- The contents should be clearly and systematically organised. At the very least, the contents flagged at the front of the thesis should coincide with the subsequent contents in actuality – that is, the page numbers and chapter titles indicated should correspond. Frequent redrafting at the preparatory stage will mean that the pagination of contents will change frequently. The contents page is often one of the last to be revised and it must not be missed.
- The citations and references should correspond exactly. The match should mean that there is no citation unreferenced, and no reference which does not appear in the main body of the text. Errors here demonstrate a certain sloppiness in the revision process, and highlight the possibility that other mistakes will appear in the main text as well as the references.

The guidelines to examiners provided by university administrators vary very little in content. The words used may differ, but they normally ask the examiner to report in five specific areas:

- originality of research;
- critical insights conveyed;
- demonstration of a capacity to conduct independent research;
- contribution to knowledge; and
- publishability of the findings.

The examiner will recognise that the thesis is a result of the collaboration between supervisor and candidate, but that the thesis is properly the work of the latter. In criticising the thesis, the examiner is inevitably criticising both candidate and supervisor, this being particularly so where methodological flaws remain that we may reasonably have expected a competent supervisor to correct.

It is instructive to examine each of the five areas above in more detail, both to clarify expectations and identify potential difficulties.

- **Originality.** The idea must be new. While it will build on the existing literature, it will embrace innovations which are new to the literature. The research must be interesting and have implications for future research or business practice. The problem under investigation must be sufficiently important to be deemed worthwhile, that is, it must pass the 'so what?' test from an unconvinced reviewer.
- **Critical insight.** The literature review is both comprehensive and current, reflecting the outcomes of previous studies, extant theory and the empirical testing of theory. The literature review should be critical, rather than just descriptive, and should not ramble around aimlessly. It should be

directed towards the development and justification of testable hypotheses. The review should be organised in a thematic manner to facilitate a critical approach and help in the identification of gaps and flaws in existing knowledge. As an examiner, all too often I read literature reviews with reference to little or nothing published in the last two years (apart from the occasional new textbook). The message is clear: the candidate completed the literature review chapter long ago and has not bothered to update it. Two areas of vulnerability often emerge here. First, in the development of theory, which is sometimes not taken seriously enough, though we know it has to be there. The ubiquitous 'agency theory' arises even though alternative competing theories are arguably more appropriate. Secondly, the jump from the literature to the hypotheses may be just too big to be adequately justified.

- **Independent capacity.** Candidates are in charge of the research project and must demonstrate that they are managing the process in a systematic, almost scientific, manner. Depending on the context of the study, the research design should ensure the internal and/or external validity of the research instrument. Failures in this regard are frequently apparent in two areas. First, candidates run out of steam around Chapter 6. All the hard work has been done, the data collected, the results analysed, but the candidates are so anxious to complete and submit that they do not do justice to their early work by skimping on the conclusions, recommendations, limitations and future research agenda. As an examiner, this is the most common criticism of theses I evaluate. Secondly, the research problem and the literature do not appear to be driving the chosen method. Frequently, candidates have preconceived ideas of their preferred method (usually a survey) even before they have specified their research idea. This must be corrected in the dissertation so that the chosen method can be seen as superior to alternatives, follows naturally from the prior literature and is consistent with the research question and hypotheses.
- **Contribution.** The study is more than just a replication. It is not the same research/instrument in a different time, country or organisation. It must be more than this. Where it relies substantially on an existing body of work, there must still be some 'wrinkle' – some significant difference – which sets it apart from existing work. As a result, we get to know something worth knowing, something non-trivial, which will make a difference. Such differences might be recommendations for change in business practice, practical business applications, new models or new theoretical relationships. However, candidates should be realistic about the limitations of their studies, and try not to overstate the significance of their findings.
- **Publishability.** If the research is new, interesting and worthwhile, then the findings will be publishable somewhere. However, as detailed in the following chapter, the research approach may restrict the number and nature of potential target journals. A well-organised dissertation, with clearly

labelled literatures and specific positive findings, will highlight publication opportunities and significantly facilitate the publication process. Where the study has clear implications for management practice, then the opportunities for publication in the professional and practitioner literatures will be increased.

Overall, examiners will be looking to make the minimum recommendation for change. They will usually not want the burden of re-examination unless they believe that major revisions must be made for a dissertation to meet an appropriate doctoral standard. The alternative examiner gradings will vary between countries, but will usually embrace the following categories:

- A – pass without further amendment.
- B – pass with only minor grammatical and editorial changes required.
- C – pass, subject to specified (minor) revisions having been made to the satisfaction of the supervisor and the university's conferring body.
- D – major revisions are required, with the revised manuscript returned to the examiner for re-evaluation within 12 months.
- E – fail, with the recommendation that a lesser award be made (e.g., of an M.Phil in a PhD examination).
- F – outright fail, with no opportunity for re-submission.

Where a *viva* is part of the examination process, this provides an invaluable opportunity for candidates to mount a vigorous oral defence of their study, demonstrating that the work is all their own, that it has been conducted in a systematic manner, and that they have a complete understanding of its intricacies. In Australia, where there is rarely a *viva* examination, problems can arise with multiple examiners and conflicting examiner opinions. But if appropriate examiners have been selected, the variance should not be great (i.e., no A and F grades to reconcile). For category D (major revisions), the nature of the required changes may vary greatly. It may mean only new literatures and major restructuring of the content – easily completed in three months. But if the examiner wants more data collection and analysis, this could be quite onerous, and for field studies may be very difficult to accomplish.

Most professional doctorates (e.g., DBAs) must satisfy very similar criteria at the examination stage. Although there are numerous DBA formats worldwide, with differing weightings attributable to the coursework and research components, the research dissertations must meet doctoral standards. They will usually be shorter than a corresponding PhD thesis (perhaps of the order of two-thirds' length) and must make a contribution to business practice, otherwise the examination criteria are very similar.

The dissertation examination may be seen as a preliminary stage in the publication process; many students will already have published during the

period of their candidature. The time period between submission of the dissertation and knowledge of the final outcome of the examination process may be both lengthy and anxious. It is an ideal time to start dismantling the thesis in order to identify publishable papers. The findings of the doctoral study can then be conveyed to a wider audience of both practitioners and academics without delay.

11

Turning Research into Publications

CHAPTER CONTENTS

This chapter targets accounting researchers at the start of their publication careers, and academics who are either just starting to contemplate a research involvement, or who need their efforts kick-started after a period of non-productivity. Those enrolled for research degrees should be aiming to publish during their candidature, and should see publication in the practitioner and academic literatures as a means of promoting themselves and their research findings. This chapter aims to answer some basic research-related questions, like why?, where?, what?, how? and which?, in a manner which at least reduces the frustrating 'trial and error' approach adopted by many inchoate researchers. The learning curve is still steep, but the gradient can be significantly reduced by observing some simple rules. The overall aim is to see more accounting students and academics conducting research, getting published, making contributions to the field and feeling the incredible buzz of seeing one's own work and name in print.

Why publish?

Research resources are spread extremely thinly around most accounting schools, especially outside the USA. Many schools may have only a handful

of active researchers, and they have the responsibility of conveying the research message to their fellow academics and doctoral candidates at a time when research output is assuming a greater prominence in all schools. The direct rewards from research involvement for both individual and academic department can be classified as:

- self-actualization;
- increased appointment and promotion opportunities;
- improved tenure likelihood;
- enhancement of teaching through the research connection;
- possible remission from teaching/administration to conduct research;
- the opportunity to win research grants;
- increased availability of consultancy assignments;
- overseas travel to present papers at prestigious conferences;
- availability to provide postgraduate research supervision; and
- attraction of resources through national university funding mechanisms (such as the Research Quantum Index in Australia or the Research Assessment Exercise in UK).

These material gains may be less important than the personal pleasure accruing from research success – the pride in one's own publications and the almost addictive effect that initial success can have on the subsequent publication record.

Where to publish?

The avenues for publishing accounting pieces are extensive and ever growing. Appendix 1 is far from comprehensive, but details 60 refereed journals in the accounting field. The initial choice is likely to lie between:

- books;
- book chapters (including case studies);
- commissioned reports;
- refereed journal articles;
- professional journal articles; and
- conference papers.

The last of these is the easiest access point, and an almost essential precursor to the publication of a refereed journal article. Most established researchers will have started their publishing careers with conference papers and professional journal articles. However, the pros and cons of each are worth consideration in more depth.

Books

Writing a book is a tempting idea. The opportunity of appearing in print, bound between hard covers, is appealing – and there may even be royalties involved. In practice, think again. The chances of writing a best-seller are remote – there is rarely enough sex and violence in accounting to produce a blockbuster, though Goldratt and Cox (1989) provide an exception. Even so, books targeted at wider non-academic audiences, like Ian Griffiths (1986) and Terry Smith (1992), have been hugely successful. Books take longer to write and rarely generate the academic kudos that an equivalent number of top refereed journal articles would. Books tend to be all-consuming, leaving little time for other research (let alone teaching) for periods in excess of six months. The review process may be as rigorous as for some refereed journals and the whole process can be very frustrating. In hindsight, many authors regret having embarked on ambitious book-writing projects, and prospective authors must not be blind to the work involved in constructing a standard undergraduate text aimed at a mass market (e.g., in management accounting a 'Horngren' for the USA (Horngren and Foster, 1991), or a 'Drury' for the UK (Drury, 1999). A specialist text is much easier. Young academics may be tempted to write adaptations of successful UK and US texts for smaller markets (for example, in Australia and New Zealand), but the dollar pay-off rarely justifies the time and energy devoted to reconstructing a tome and generating culturally appropriate support materials, at an important time in their careers. Having said that, at least one book, or major contribution to a book, like the successful competitive research grant application, contributes to the rounded resumé so sought by potential employees. However, it is probably not the best place to start; better to have a number of refereed and professional journal publications under one's belt before even contemplating writing a book.

Book chapters

These provide a much easier alternative, and embrace the contribution of case studies to an edited collection. However, you will still have to learn the art of compromise because the publisher will have at least one eye on the target market, while the editor must ensure that the whole volume conforms to a particular house style – usually his or her own! The result is that any contributor will be required to change the actual words and phrases they have chosen and be forced to add or delete paragraphs which they may believe to be unnecessary or essential, respectively. Many multi-author books are instigated in single university departments, but while all contributors are good friends at the outset, they frequently are not by the time the volume reaches the bookshops.

Commissioned reports

Once you have developed an area of recognised expertise, commonly associated with a doctoral research topic, opportunities will arise to produce (usually for money) commissioned research reports. The borderline between this activity and consulting is extremely blurred, especially if a niche of expertise is such that it results in a plethora of reports in a similar vein but slightly different environments (e.g., the same thing in different organisations, different industries, different countries, etc.). However, many 'research streams' in the academic literature make certain individuals subject to the same criticisms. Reports of this nature are unlikely to carry a great deal of weight with academic peers, but they may foster valuable industrial networks and help to build a reputation within the profession.

Refereed journal articles

These are generally regarded as the pinnacle of achievement for an academic. Certainly for promotion and tenure purposes, the quantity and quality of your contributions to refereed journals will be important, if not paramount, to your success. Parker et al. (1998, p. 381) add weight to this perception with quotations from accounting professors responsible for academic recruitment. In assessing the relative importance of research, teaching, administration and community service in academic appointment and promotion, one interviewee reveals that: 'Publications is 101 per cent. Everything else is zero.'

The downside is that the better journals are difficult to get into – they reject something like 70% of pieces submitted (for *The Accounting Review* this figure is over 90%). The other, related, problem is the length of time it takes to get an article published in a refereed journal. To understand this delay, consider the nature of the reviewing process:

The relationship between author and referee(s) allows alternative forms of review process to emerge:

- open – they (referees) know who you (author) are *and* you know who they are;
- single-blind – they know who you are *but* you do not know who they are; or
- double-blind – referees and author(s) are anonymous to each other.

The best refereed journals use the double-blind method, but that does not totally remove the potential for reviewer-bias because both topic area and writing style may reveal the identity of the author. It is also quite likely that the reviewer has already seen an earlier version of the paper, in conference paper or working paper form, a factor that the author should aim to exploit to his or her advantage.

The referees generally determine the publishability of the paper and they will report back (separately) to the editor, giving an opinion on the paper. This generally takes between three and six months, sometimes longer. Where the referees disagree, the editor may resort to a third referee as an arbiter, extending the process by a further two to three months. When the response from the editor finally arrives it will come in one of the following alternative forms:

- Editorial rejection: returned as being unsuitable without recourse to referees. If the target journal is inappropriate, the editor may suggest an alternative which is more suitable. In the worst scenario, the editor may consider that the paper is so badly written it is unfit for refereeing.
- Rejection with referee reports: short referee reports are not good since they convey the message that there is no hope for the paper.
- Rejection but with rewrite suggestions: the referee reports are constructive in detailing the reworking that is needed to make the paper adequate. These revisions may be so excessive that they constitute a new paper (e.g., the redesign of an experiment or conducting a fresh sample), but at least they convey the promise that there is a paper in there somewhere, even though it may never be acceptable to this particular journal.
- Yes/but acceptance: the referees like the paper but have identified serious flaws which preclude its immediate acceptance. However, this should be treated as an 'accept', even if the corrections are extensive and time-consuming, because addressing them should guarantee success, as long as the journal editor remains consistent in his or her approach.
- Acceptance with minor revisions: the referees feel compelled to make some changes (this is part of their job after all), but they are so minor that they can often be completed by return of post. Success!
- Acceptance without revision: the referees recognise your genius and demand your publication!

The last event is extremely rare, and virtually unknown in the top journals. Several iterations of the reviewing process may take place (usually two or three) before the referees are finally satisfied. These iterations can prove a tortuous process, especially when the referees are looking to push the paper in seemingly opposite directions. Clear guidance from the editor as to their priorities is most helpful in such circumstances. A year will normally have elapsed between submission of a paper and its acceptance, at least another year will pass before it appears in print. On average, two years work on researching the project and workshopping the paper will have been spent before its submission to the reviewing process. That is, four years in total between starting work on the paper and seeing your name in lights. It may be longer if one of the referees dies or changes his or her mind during the review process. Herein lies the major drawback of the refereed journal article – any one has a lengthy life-cycle, making it essential that, as an author, you have several papers undergoing the review process simultaneously. It is not advisable to try to speed up the process by submitting the same paper to different journals at the same

time. In fact, the better journals have a policy specifically prohibiting the acceptance of articles treated in this way.

The time-lags in the refereeing and printing process usually go beyond that required to guarantee integrity of content, and most journal editors know it. They are conscious of the threats to relevance from lack of timeliness, but they have a difficult job when dealing with (usually) unpaid referees trying to fit an often demanding, but discretionary, activity into their busy schedules. Electronic submission of papers for consideration and online publishing should help to reduce some of the time lags, but most journals have been extremely slow in adopting procedural changes.

Professional journal articles

All of the disadvantages of refereed articles are magically reversed when contemplating the professional journals:

- The articles do not take as long to write: more than a week's work spent on a professional piece is generally considered excessive.
- The content does not have to be earth-shattering or even make an original contribution to the literature. It does have to be relevant and timely, and be able to make practitioners aware of issues they should know about, or may have forgotten.
- The review process is speedy: the editor may accept the piece without recourse to others, but often they will seek 'expert' opinion.
- Once accepted (and acceptance without revision is common) the article will usually be published within three months, providing that it can be accommodated within the backlog and the advertising pages. The more pressing the issues examined, the quicker the article appears in print.
- Feedback from readers will be quick (positive or negative) and is much more common than that associated with refereed journal articles.

Articles for professional journals are relatively easy to write and may be a direct consequence of teaching experiences. They can make active contributions to the professional development of accounting practitioners by bringing to their attention the results of relevant academic research, the potential for applying new methods, or for applying old methods in new ways. For example, articles on CAPM, product life-cycle and time-series analysis frequently appear in the professional literature because they have been written with respect to a particular issue and 'packaged' to say something new. They do not need to be academically rigorous and a shortage of citations is a distinct advantage. They do need to be brief, clear and concise. Catchy titles and acronyms are helpful too, giving the editor an ideal selling point. For academics who have published little or nothing, but have concentrated their efforts on teaching and assessment, this is the easiest means of penetrating the research genre. However, do be prepared to see versions of papers in professional

journals that differ vastly from the ones which were actually submitted and accepted. Editors take the word 'editing' very seriously, and you can expect to see both the title changed and the content of 'your' paper abbreviated to fit the space requirements of the journal, often without reference back to the author. The 'dumbing down' of the content of practitioner journals in recent years is an unfortunate but unavoidable observation. Their magazine-like format, in which pieces on cars and holidays compete for space with technical issues, has a dual effect, which threatens the informing process:

- publishers are less willing to publish academic pieces, preferring shorter contributions from consultants;
- academics are less willing to submit pieces to the practitioner literature because they perceive the medium to have been devalued.

The *Harvard Business Review* is worthy of specific mention here, as a type of journal which really fits neither of the above descriptions for refereed or professional journals. This is especially appropriate given that it is a prestigious, highly-cited journal that attracts contributions from the biggest names in academic accounting. The content, however, is rarely academically rigorous. The focus is on findings and the application of findings to practice, rather than on methodology or theoretical underpinning. The editorial process is very different from that for refereed journals too. Indeed, once it is convinced of the validity of an idea, the editorial team at *HBR* will work closely with the authors in the actual writing of the article to ensure that it reaches the target practitioner audience appropriately. Traditional accounting academics may accordingly begrudge the reputation that *HBR* has, but it remains a very useful citation to have in one's resumé.

Conference papers

The conference paper is rarely an end in itself in the accounting discipline, but may find itself in a published collection of conference proceedings. In other disciplines, notably marketing or information systems, published conference proceedings are more often the final resting place for a paper. The refereeing process for accounting conferences varies considerably, and this is reflected in the content that is considered acceptable: many conferences will review abstracts only, and make a decision on the acceptability of the whole paper based on preliminary evidence – at a stage when the paper itself may be in an early draft. Such conferences provide excellent vehicles for exposing preliminary findings. Where conference convenors wish to review the whole finished paper prior to deeming it acceptable for presentation, in a manner close to that for refereed journals, they seem much more inclined to accept polished papers, almost complete and ready for publication, on the grounds that internal and inter-university workshopping should have already been used to iron out flaws of construction and presentation. The proliferation of accounting

conferences in recent years, where the organisers recognise that potential delegates must be presenting for their attendance to be funded, means that it is no longer difficult having a paper accepted for presentation at an international conference.

What to publish?

There are few restrictions on the topics for professional journals, or the manner in which they are written, as long as they comply with tight space restrictions. The same cannot be said of refereed academic journals, where the topics deemed appropriate and the approach deemed acceptable often seem to be unduly, even absurdly, constrained. These issues are explored in depth by Gray (1996).

Peat, Marwick & Mitchell's programme, *Research Opportunities in Auditing* (1976), identified a number of characteristics of research problems that need to be addressed for them to be attractive to the academic community:

- they should be interesting;
- they should be capable of scientific research, in terms of testable hypotheses and discernible outcomes;
- they should be rooted in knowledge and research methodologies which are easily attainable;
- datasets and sample information must exist to allow the development of theories and hypotheses;
- the problem should be sufficiently important to attract research funding; and
- the likely outcomes must be of a standard to justify publication in a refereed journal or another respected publication.

These points provide useful guidance, but in spotting a specific research question there is no substitute for reading the research literature. Keeping up with the current research literature across a wide span of interests is extremely time-consuming – some would argue impossible – if you are trying to *write* as well as *read*. There are short-cuts:

- make use of online alerting services, based on both journal title and key word, so that you are well informed of the contents of all relevant recently published journal articles;
- make use of online access to working paper series so you can see what will appear in the journals at least a year in advance of appearance;

- read the 'abstract' of the paper – you may have no choice because that may be all that is available online;
- when full-text or hardcopy is readily available, begin by reading just the beginning and end of papers:
  - the 'abstract' and 'introduction' to see if the content and the approach are of interest; and
  - the 'conclusions' and 'limitations' to see what the paper has *not* done.

  These will frequently identify flaws in methodology that can be corrected by replication with 'a wrinkle' and opportunities for future research;
- read anthologies of the current state of research which will identify those fields still to be explored or which have been researched inadequately;
- attend conferences to see what other people are working on – their ideas can often be adapted quite successfully without plagiarism.

In the accounting and finance areas the research topics can be grouped broadly as follows:

Analytical and non-empirical areas

- Formal, highly mathematical expositions based on the information economics and utility theory literature (e.g., in the *Journal of Accounting Research*).
- A return to fundamentals with a focus on 'measurement' or the 'value' of accounting information for decision-making.
- Critical theory research.

Focused empiricism

- A move away from narrow studies devoted to the 'information content' aspects of stock price reaction towards behavioural aspects, often in laboratory-type conditions.
- Applications in particular (under-researched) institutions, like insurance, healthcare, the public sector and non-profit organisations, through survey-based or archival studies.

Socio-political structure

- Case-based development of systematic bodies of knowledge regarding accounting in complex situations.
- Unstructured, positive, exploratory, case-based investigations of actual practice and, for example, the impact of modern manufacturing technologies.
- Field-based experimental studies, which are still extremely rare in the literature.

Hypothetical deduction

- Survey-based statistical inference to appraise the effect on accounting systems of contingent variables.
- Impact of changes in accounting systems over time relative to movements in competition and strategic direction.

All of these topics may be described by the NIRD acronym (i.e., new, interesting, reproducible and defensible), however, not *all* approaches are equally acceptable to *all* journals. Case study material, for example, is still difficult, though decreasingly so, to get published (perhaps because it fails the 'reproducible' test). It would therefore be a waste of time to send such a study to the *Journal of Accounting Research*, for example. Similarly, in Chapter 9 we noted the difficulties that authors experience in publishing critical theory research in other than two or three of the top accounting journals. Reading the research literature involves 'researching the editors' so that you know what they have published and where, and the style and topics likely to be acceptable to them as editors. In this way journal submissions can be targeted towards journals that increase their likelihood of acceptance.

A paper is more likely to be accepted if it conforms to a standard structure that demonstrates the adoption of a systematic scientific approach. A paper would normally be developed through the five distinct stages detailed in Figure 2.2. These would be supported by an 'abstract' providing an overview of the paper. Many authors neglect the importance of this 'sixth' stage by using it as a receptacle for sentences and paragraphs used elsewhere in the main text. This is a big mistake. The abstract is the first thing the editor/reviewer reads and great care is necessary to ensure that it 'sells' the paper to the reader, while spelling out what it does and why it is important. It should be possible to identify the nature of the key results of the paper and the contribution it makes to the literature from the abstract and the initial pages of the introduction. If the abstract is poorly written and unexciting, the editor may read no further! The write-up of the five stages demands closer attention.

Research problem

The research question must be specified precisely, and in a manner which details the motivation for its study. It must be a problem worth studying. The importance of this section cannot be overstated; it provides the opportunity of 'selling' an idea and of setting out what the paper does clearly so that the key points are not buried in a mass of trivia. The paper must pass the 'so what?' test. (See also pages 20–3.)

Theory and literature review

The literature review must be current and fairly reflect both relevant theory and the outcomes of previous studies. It should provide a critique of earlier

work, pinpointing flaws in the approach of others. Most importantly, it must not ramble about in an apparently undirected manner; it should be precise and link to the development of hypotheses. (See also pages 23–4.)

Hypotheses

Hypotheses must be driven by underpinning theory, and the latter should be sufficient to avoid making speculative leaps from theory to hypotheses. A single paper would not normally contain more than seven separate hypotheses – more than this and it starts to look like another paper. (See also pages 52–3.)

Research methodology

The method of investigation chosen should be consistent with both the research question and the hypotheses. The experimental design should ensure the internal and external validity of the research instrument is, as far as possible, in the context of the study. The method of investigation should be justified (e.g., why a survey?) and must be structured in a scientific manner to demonstrate that it will test the question being researched. Any sampling method used should be described, justified, and shown to be scientific in its selection. (See also pages 53–4.)

Results, conclusions and discussion

The empirical analysis should be appropriate to the chosen method and should be directed towards the measurement and testing of the hypotheses. The results should be clear, avoiding unnecessary mathematical complexity and 'mini-lectures' on the theory underpinning the application of standard tests. The author(s) should be realistic about the limitations of the study and not overstate the significance of its conclusions. Areas where the study might be extended and related research areas should be highlighted.

The paper should conclude with a set of references which exactly coincide with the contents (i.e., no citation is unreferenced and no reference is uncited). A checking technique used by many referees (including this author) to verify the care which authors take in completing their papers is to start their review by ensuring that the equation 'citations = references' balances. Lack of care by authors in this regard is a fair indication of their likely sloppiness and lack of attention to detail in other aspects of their paper. (See also pages 160–1.)

The whole paper should be interesting, readable and clear, reflecting the motivation of the author(s) to convey a message which they feel should be heard. A length of 25–30 pages, including appendices, is the norm, but both longer and shorter papers can be justified. If in doubt about including paragraphs,

tables or appendices, then they should be omitted. The same applies to footnotes (or endnotes, depending on the journal), the number of which should be kept to a minimum. A check of the typical article in the target journal will quickly show the elements of house style and what editors are looking for. Make sure that articles submitted to journals for publication comply with that journal's editorial style in every respect – and particularly in the method of referencing. If you use the style of another journal, then you may send an unwitting message to the referee that this paper was originally destined elsewhere but has already been rejected by that journal! The use of the 'endnotes' software fortunately now makes the once time-consuming and tedious reformatting of references a simpler task.

How to publish?

There are many conflicting views on what constitutes an appropriate and ethical method of answering the 'publishing' question. They boil down to 'knowing your editor' and 'respecting' the views of referees. Some authors subscribe to the view that the submission of three-quarter finished papers to journals is justified by the fact that referees are bound to want to make changes (that is their job after all) and why not let them finish the paper for you? It is difficult to support this view. Apart from the ethics of the process, such practice will soon get you a bad reputation with both editors and reviewers. Generally, it is good practice to submit a polished piece to a journal, one which you would be happy to see published as it is, but which may still have some 'rough edges' that reviewers can further refine. A useful system would be:

- Write a rough first draft yourself.
- Edit the first draft with clarity, sense, structure and order of materials in mind, ensuring that spelling, grammar and references are perfect.
- Ask somebody else in the field, whose views you respect, to look critically at the paper and encourage his or her use of the red pen. Reading and altering the first drafts of others is an extremely time-consuming operation if done properly, so care should be taken to ensure that you, at least, have made a reasonable first attempt, and that you do not pepper the same colleague with too many first-drafts per annum – not if you wish to retain their friendship and co-operation!
- Revise the paper with these views in mind. Some criticisms may be so fundamental as to necessitate a complete rethink/rewrite.
- 'Workshop' the paper at one or two other universities, to seek the views of a wider (possibly interdisciplinary) audience.
- Submit the paper to a respected conference in the field. This may be a regional, national or international conference, depending on timing.

- Further refine the paper based on these comments and submit this version to the target journal.

Editors expect authors to adopt a systematic approach, so it is worth reminding them that this has been undertaken by making acknowledgements on the front page of the paper to individuals whose comments have been sought and to the participants of workshops, seminars and conferences where the paper has been read. A variation on this methodology that is frequently adopted in practice is to send a polished first draft of the paper to a member of the editorial board of the target journal for comment. Provided that this person's comments are then incorporated in the paper, the inclusion of his or her name on the front cover should elicit a positive reaction from the editor and may even induce them to use the same person as a formal reviewer.

In targeting the journal in the first place, check out the editor and members of the editorial board. What are their areas of strength? Have they published anything in the subject areas of your paper? This approach allows you to formulate a short-list of likely referees (though journals will commonly also use *ad hoc* reviewers from beyond the editorial board too). It should also highlight pieces in the literature which it would be prudent to cite, and those which it would be rude or even foolish to ignore.

When submitting the paper write a brief accompanying letter to the editor, specifying clearly why you think this is the appropriate journal to publish this paper, and the contribution that it makes to the current debate in the area.

When your paper has entered the reviewing process, be patient. Try not to call editors to check on the paper's progress as this is annoying and will soon get you a bad reputation. It is reasonable, though, to write a letter of enquiry or send an e-mail after a period of three or four months has elapsed without response. When the response arrives, it may indicate acceptance, although, more likely, rejection. However, a rejection is often not what it seems, as noted earlier. The initial reaction to rejection (of any description) and negative reviewer reports is usually swearing, and perhaps tears. This is normal, and an essential part of becoming a productive publisher is growing a hide thick enough to cope with repeated rejection. One immediate course of action, advocated by many, is to despatch the paper to the next-choice journal by return of post. This is rarely an optimum strategy, especially where it is highly likely that the *same* referee will be used by *different* journals – the narrower the topic area, the higher the probability of this happening. There is nothing that annoys a reviewer more than to be asked to referee a paper that he or she has already refereed and in which none of the previously recommended changes have been made. Such action can again cause damage to your long-term reputation as a researcher.

It is much better to consign the rejected article to a drawer, forget about it for a while and concentrate on another project. A sensible researcher will always have at least four projects running simultaneously, and all at different stages of the research process, in order to generate a continuous flow of output. Return to the rejected paper after a week or two and re-read the referee

reports carefully. If the tenor of their remarks still causes a state of excitement and agitation, then re-consign the paper to the drawer for a little longer, at least until you have calmed down sufficiently for teeth to become ungritted. When the referee reports can be read in a relatively unemotional manner, note the thrust of their comments and re-read your paper to determine the extent to which they are justified. It is rare not to come away feeling 'the referee has a point there' in at least one or two instances. The referee reports can then be used positively and constructively in the rewrite of the paper:

- Specify the points made by each reviewer (usually two) separately and classify them as major (methodology, missing literature) and minor (labelling, positioning, footnotes) criticisms. Reviewers should normally be concerned with some common issues.
- Address each of the specified points. This does not necessarily mean agreeing with the reviewer on every point. If there are arguments for *not* making the suggested changes, then articulate them (reviewers *do* make statistical errors, and it is unlikely that they are as familiar with the current literature in a narrow specialist area as the author).
- Do not attempt to make changes which are merely 'window dressing' – they fool nobody. Major criticisms will call for substantive reworks, possibly involving the results of new experiments or the adoption of improved testing procedures.

If all of the major criticisms can be addressed and/or corrected, then the revised article can be resubmitted to the same journal. If there are major gaps remaining, where the paper is still vulnerable, it may be better to target a lower-tier journal.

When resubmitting, the accompanying letter to the editor should thank the editorial team for their constructive remarks and identify the major changes made to improve the paper. Separate reports for each reviewer, detailing how each point has been addressed, and where in the revised paper changes to the text have been made, should be attached. The editor will normally return the resubmitted paper to the original reviewers, together with your comments on the points raised. They may be satisfied and recommend publication, or they may ask for more (on the same points) or identify different ones.

Reviewers may be obstinate or even hostile, or worse, they may change their minds. It is reasonable to make a fuss and/or ask for a different reviewer in such circumstances, but in my experience this is not normally helpful. Usually the editor has made the decision that *this* paper will not be published by *this* journal, and in such circumstances it is advisable to look elsewhere.

Concluding remarks

Appendix 1 illustrates the range of publications available in the accounting area, including refereed, non-refereed and some offering a blurred distinction between the two. Some academics (e.g., Beattie and Ryan, 1989; Brin et al., 1996; Brown, 1996; Hull and Wright, 1990; Parker et al., 1998) have addressed the quality issue and/or attempted scientific ranking methods for journal reputation. The ranking of Appendix 1, tiered in a manner similar to league tables – a Top 20, followed by a Next 40 – is a subjective, personal listing, though inevitably influenced by the views of colleagues. This ranking will change over time because some journals will significantly 'lift their game', while others will deteriorate in quality, often as a result of a change in editor.

For those seeking academic appointments or promotions, a combination of both refereed and professional journals is helpful, but with a clear emphasis on the refereed journals. In this way you can demonstrate an ability to relate to both the profession and fellow academics. Premier division publications will be particularly well regarded if they are representative of a stream of research productivity, while the resumé should include both sole-author and multiple-author publications to demonstrate that you can work alone and as part of a team.

This chapter spells out the 'rules of the game' based on personal experience. If it facilitates one publication that would otherwise have gone unwritten, or inspires one new researcher, then it will have served its purpose.

Appendix 1: Ranking of Accounting Journals

Premier Division (Top 20)

Abacus
Accounting and Business Research
Accounting and Finance
Accounting, Auditing & Accountability Journal
Accounting Horizons
Accounting Organizations and Society
Administrative Science Quarterly
Behavioral Research in Accounting
British Accounting Review
Contemporary Accounting Research
Critical Perspectives on Accounting
Journal of Accounting and Economics
Journal of Accounting Research
Journal of Banking and Finance
Journal of Business Ethics
Journal of Business Finance and Accounting
Journal of Business Research
Journal of Management Accounting Research
Management Accounting Research
The Accounting Review

First Division (Next 40)

Accounting, Accountability and Performance
Accounting, Business and Financial History
Accounting Education: An International Journal
Accounting Educators Journal
Accounting Forum
Accounting, Management and Information Technologies
Accounting Research Journal
Advances in Accounting

Advances in International Accounting
Asia Pacific Journal of Accounting
Asian Review of Accounting
Auditing: A Journal of Practice and Theory
Australian Accounting Review
Australian Journal of Management
British Journal of Management
Decision Sciences
European Accounting Review
Financial Accountability and Management
Harvard Business Review
Issues in Accounting Education
Journal of Accounting and Public Policy
Journal of Accounting, Auditing and Finance
Journal of Accounting Case Research
Journal of Accounting Cost Research
Journal of Accounting Education
Journal of Accounting Ethics
Journal of Accounting Literature
Journal of Applied Accounting Research
Journal of Applied Management Accounting Research
Journal of Cost Management
Journal of International Accounting, Auditing and Taxation
Journal of International Financial Management and Accounting
Managerial Auditing Journal
Managerial Finance
Pacific Accounting Review
Research in Accounting Regulation
Review of Accounting Studies
Review of Quantitative Finance and Accounting
The Accounting Historians Journal
The International Journal of Accounting Education and Research

Appendix 2: Sample Paper (1)

This paper was originally published in *Managerial Auditing Journal*, Vol. 16, No. 1, 2001, pp. 40–9. It is reproduced here by kind permission of MCB Press, UK.

Structure versus Judgement in the Audit Process: A Test of Kinney's Classification

Malcolm Smith*, Brenton Fielder*,
Bruce Brown* and Joanne Kestel**

*University of South Australia,
**Murdoch University

Keywords: audit, judgement, creative accounting, Big 6, structure, accounting policy, Kinney, Big 8

The authors acknowledge the helpful comments of Professors Keith Houghton, Gary Monroe and Brenda Porter, and those of delegates to BAA Conference, Manchester, April 1998.

Abstract

Sullivan (1984) suggests that the alternative audit approaches adopted by accounting firms be expressed in terms of 'structure' and 'judgement', with a division provided by the degree to which auditor judgement is replaced by structured quantitative algorithms. Cushing and Loebbecke (1986) attempt to operationalise this division by examining the guidance provided to practising auditors by their firms. Kinney (1986) extends this study by classifying accounting firms as 'structured', 'intermediate' or 'unstructured' in terms of their audit methodologies.

This study provides a test of Kinney's classification by examining the tolerance of accounting firms to accounting policy choices which have an income effect in their clients' financial statements. The paper argues that those firms with a structured audit approach will manage audit risk through a greater reliance on mechanistic procedures, resulting in a greater tolerance of income manipulation. The results are confirmatory for the period under study, but evidence is provided to suggest that audit firms have subsequently become less diversified in their approach.

1. Background

Organisational theory (e.g., Burns and Stalker, 1961; Mintzberg, 1979) has suggested the 'machine' and the 'organism' as analogies forming a basis for refined reasoning. In auditing, these analogies have been discussed in terms of the concepts of 'structure' and 'judgement' (Dirsmith and Haskins, 1991).

Auditing has variously been regarded as a well-structured and mechanistic process (e.g., Joyce and Libby, 1982) or as a judgemental process in which the audit is client dependent (e.g., Dirsmith and McAllister, 1982). Stringer (1981), among others, observes the trend towards increasing structure in auditing decision-making with the use of quantitative methods and well-documented procedures. Sullivan (1984) highlights the two camps into which auditors fall:

- those who favour structured quantitative algorithms over auditor judgement, and
- those who believe that such quantification is always unjustified because considerable professional judgement will always be required.

Cushing and Loebbecke (1986) explore this distinction with an empirical study of the guidance provided by accounting firms to their practising auditors. Their study of the policy manuals of twelve large public accounting firms revealed dramatic differences between firms in terms of the degree of 'structure' apparent in their audit methodologies, defining 'structure' as 'a systematic approach to auditing characterised by a prescribed, logical sequence of procedures, decisions and documentation steps, and by a comprehensive and integrated set of audit policies' (p. 32).

Cushing and Loebbecke noted that all firms placed a good deal of emphasis on pre-engagement planning and internal control questionnaires, but that beyond that they might be categorised as highly structured, semi-structured, partially structured and unstructured, with the extreme positions characterised by, respectively:

- quantification of audit risk; detailed comprehensive guidance; shift of audit decision-making from the auditor to the central firm, and
- no specification of the level of detail, integration or quantification.

Cushing and Loebbecke (1986) recommend that future research be directed towards identifying the differences in firms associated with structure and the consequent impact of alternative audit approaches. This recommendation provides a motivation for this study.

Kinney (1986) extends the work of Cushing and Loebbecke (1986), noting that the unstructured approach is associated with more judgement considerations being left in the hands of the field auditor. Kinney uses the results of an independent survey, together with those from the Cushing and Loebbecke study, to classify 22 auditing firms (the, then, 'Big 8' and 14 smaller firms) as follows:

- structured
  - Deloitte, Haskins and Sells (DHS)
  - Peat, Marwick, Mitchell (PMM)
  - Touche Ross (TR)
  - two non-Big 8 firms
- intermediate
  - Arthur Andersen (AA)
  - Arthur Young (AY)
  - Ernst & Whinney (EW)
  - three non-Big 8 firms
- unstructured
  - Coopers & Lybrand (CL)
  - Price Waterhouse (PW)
  - nine non-Big 8 firms

It is this classification which forms the basis of the test conducted in this study. The degree of audit structure has been found to be associated with the financial disclosure patterns of clients. Morris and Nichols (1988) show that structured firms are more consistent in their treatment of accounting principle consistency exceptions; Williams and Dirsmith (1988) show that structured firms are more timely in their release of client financial statement disclosures. This study extends this area of research by examining the impact the degree of audit structure has in individual firms on the tolerance of income increasing/reducing accounting policy choices among client companies. This paper argues that audit structure impacts on such tolerance via perceptions of audit risk, the risk of incorrectly attesting that a client's financial statements are true and fair.

Dirsmith and Haskins (1991) note that audit risk as a planning construct is receiving increasing attention in the literature (e.g., Fellingham and Newman, 1985) and that high degrees of audit risk are associated with increased evidence gathering to support the audit opinion (e.g., Graham, 1985).

Contemporary auditing standards and the literature (e.g., Graham, 1985; Dirsmith and Haskins, 1991) recognise that internal control risk and inherent

risk are interdependent and must be considered together in planning an audit so as to determine the desired detection risk. It has been suggested that audit structure may impact the assessment of inherent risk, whereby a more thorough evaluation of all the important quantitative variables will produce consistent auditor judgements (e.g., Joyce and Libby, 1982). Sullivan (1984) puts forward the opposing view, by suggesting that financial reporting requirements are too complex to be represented satisfactorily by quantitative measures alone, and that informed auditor judgement will always be required.

The response of audit firms to the ambiguity of approaches to inherent risk assessment suggests that 'audit firms which vary in terms of structure would orient differently to such an assessment' (Dirsmith and Haskins, 1991, p. 75). Dirsmith and Haskins conclude that researchers can usefully study auditing with reference to the public accounting firms' underlying root metaphors and world theories. Their study focused primarily on differences relevant to the assessment of audit risk using the 'mechanistic world' and 'organic world' hypotheses. The mechanistic world hypothesis sees auditing as a structured process that emphasises parts, priority relations within the parts, and the dominance of quantitative versus qualitative components of the audit judgement. Alternatively, the organic world hypothesis views auditing as a judgemental process emphasising holistic integration with more qualitative considerations forming part of the judgement process.

Dirsmith and Haskins postulate that:

> ... auditors' perceptions of inherent risk assessment, as well as the language they use to describe this assessment for specific clients, may be influenced by the world theory subscribed to their respective audit firms. (1991, p. 75)

Further, they state that mechanistic, structured audit firms would tend to discount their focus in audit areas that are qualitative in nature and less subject to analytic evaluation. Accordingly, such firms would be likely to focus on those parts of the audit that are 'relatively structured, programmable, concrete and familiar ...'. Conversely, less structured firms are perceived to have a more balanced focus on both quantitative and qualitative forms of evidence.

These hypotheses confirm a nexus between structure of the firm and the attitude toward risk assessment. We perceive that auditors in structured firms place more reliance on their relative sophistication in, for example, outcomes of analytical review strategies (including analysis of quantitative non-financial indicators), sampling methodologies and greater strategic focus in the global audit approach. We perceive, therefore, that structured firms, while recognising the relative importance of assessment of both qualitative and quantitative risk factors in planning and conducting an audit, are able to reduce the emphasis on qualitative assessments due to their reliance on identifying risk factors using strategic quantitative analysis. It should be recognised that structured firms deploy substantial resources into technical divisions that produce high-quality generic research and technical data for

use by audit field staff (e.g., industry statistics, generic qualitative industry risk assessments and programs, contemporary technical issue papers, circulars and so on). We perceive the availability of such data is significant in structured audit firms' assessment of the overall risk involved in a client. It is this reliance which leads to the proposition that structured firms may be more tolerant of accounting choices selected by audit clients for the purpose of income 'smoothing' or 'manipulation'. These firms have resources that impact their decision-making about the overall audit risk and ramifications of offering an inappropriate audit opinion. It is not suggested that the fundamental audit approach of structured firms is flawed, but the focus of structured firms seems to be more on the longer-term view of audit risk of client failure and short-term tolerance of income manipulation.[1]

In order to reach these same conclusions, it is contended that unstructured firms require a greater level of investigative qualitative assessment, and may be less tolerant of income manipulation by having access to more reliable qualitative data. Cushing and Loebbecke (1986) confirm a correlation between highly structured firms and reduced opportunities to apply professional judgement. We hypothesise that greater reliance on features of the audit firm structure (including detailed audit manuals, procedures and strategies) narrows the relative depth of qualitative assessment and broadens the tolerance to income manipulation ('income smoothing') perceived by the firm as non-threatening to audit risk. This proposition is tested by exploring the degree to which the clients of Big 8 audit firms (classified according to Kinney, 1986) make accounting policy choices which impact on income. The circumstance not controlled by this experiment is the nature of the audit client portfolio, as certain audit firms attract clients that engage in certain accounting policy settings.

2. Research method

Annual reports of all 463 West Australian public companies were examined for financial years ending 1987 and 1988 to determine the incidence of accounting policy change. Those companies, numbering 96 in all, with no 1987 and/or 1988 accounts available, either because of incorporation post 30/6/87, failure prior to 30/6/88, or missing data, have necessarily been eliminated from the study. The financial years under study corresponded with the publication of Kinney's classification and provide an opportunity to

1 Here we follow developments in positive theories of accounting choice (e.g., Mian and Smith, 1990; Anderson and Zimmer, 1992a) suggesting that variation in accounting methods reflects the firm's demand for efficient contracting, is a function of differences in firm circumstances, and reflects the desire of firms not to make accounting changes which will reduce future operating profits.

investigate activities of Big 8 firms immediately prior to a series of mergers that reduced the numbers of the major companies.

A change in accounting policy from one financial year to the next was defined as a change in disclosed policy choice.[2] Although mandatory changes were identified, only the effects of discretionary changes were examined. These changes were determined by reference to the auditors' report and to the Notes to the Accounts (and in particular the note describing Significant Accounting Policies required by Australian Accounting Standard AASB 1001).

Every disclosed accounting change by the firms under observation was treated as an independent case, and data on all changes was collected, regardless of impact. After analysis, the changes were classified into five groups:

- change in response to a qualified audit report;
- mandatory changes in response to legislation and new/revised accounting standards;
- changes with indeterminate income effect (even though a clear balance sheet impact may be apparent);
- income increasing changes, including changes relating to normal and abnormal operations and changes which resulted in expenses being treated as extraordinary items even though they might reasonably have been included as normal/abnormal;
- income reducing changes resulting in reduced after tax earnings.

Comparison of the independent assessments of accounting policy change made by the investigators resulted in substantial agreement of classification. A complete reclassification undertaken by the investigators at a three-month interval was substantially confirmatory, with only four per cent of changes being reclassified.

For all companies in the dataset the following information was also collected:

- auditor (where a change of auditor had occurred over the period that observation was removed from the population);
- status (defined as 'failed', including 'failing', or 'non-failed');
- size, measured by total assets; and
- industry group.

The information was gathered to test the possible impact of these variables on the sample results. It was considered that the hypothesised results would be strengthened if the allowance of changes to accounting policies is not identifiably linked to auditor change, nor influenced by the financial condition, size or industry of the company being sampled.

2 In so doing we follow Anderson and Zimmer (1992b, p. 58) in suggesting that 'accounting techniques are "independently" chosen each year, dependent on firm circumstances'.

FIGURE 1

Total accounting policy changes	No. of companies	Mandatory changes	No. of companies	Discretionary changes	No. of companies
0	191	0	276	0	249
1	105	1	73	1	77
2	48	2	18	2	32
3	17			3	8
4	4			4	1
5	2				
Total companies	367	Total companies	367	Total companies	367
Total changes	278	Total changes	109	Total changes	169

Distribution of accounting policy changes across companies

3. Results

The analysis of the population of the 367 companies for which complete data was available revealed that 176 companies made accounting policy changes, resulting in a total of 278 changes, as follows:

Response to audit qualification	9
Mandatory change	109
Indeterminate income effect	52
Income increasing	79
Income reducing	29
	278

The accounting policy changes were distributed across companies in accordance with Figure 1.

Of the nine companies with three or more discretionary accounting policy changes, seven were audited by Big 8 companies, five were from the extractive industry, only one was a 'large' company and four 'small' as designated in Figure 4. Of the 28 changes made, half were income increasing.

The detailed figures in Figure 2 enable a number of statistical evaluations to be made concerning the association of choice of auditor with the distribution of accounting policy changes. The tests described below in Figures 2a, …, 2d are all based on data drawn from Figure 2.

The Big 8/non-Big 8 split is a significant factor in determining the incidence of accounting policy changes. Figure 2a gives $P < .002$ when all changes are considered, while Figure 2b yields $P < .01$ when mandatory changes are excluded.

FIGURE 2

		COMPANIES		ACCOUNTING POLICY CHANGES						
AUDITOR	NO. OF COMPANIES	No changes	Policy changes	Mandatory	Response to AQ	Neutral	INCOME INCREASING		INCOME REDUCING	TOTAL
							Above line	Below line		
(DHS) DELOITTES	15	6	9	7	1	3	1	3	2	17
(PMH) PEAT MARWICK	18	9	9	3	0	3	2	3	2	13
(TR) TOUCHE ROSS	26	8	18	11	0	2	8	9	2	32
(AA) ARTHUR ANDERSEN	43	18	25	11	4	5	9	4	5	38
(AY) ARTHUR YOUNG	38	19	19	9	0	5	7	0	3	24
(EW) ERNST & WHINNEY	32	15	17	11	0	6	7	6	2	32
(CL) COOPERS & LYBRAND	25	13	12	11	1	2	1	0	1	16
(PW) PRICE WATERHOUSE	18	9	9	8	0	4	0	0	1	13
BIG '8'	215	97	118	71	6	30	35	25	18	185
NON-BIG '8'	152	94	58	38	3	22	8	11	11	93
TOTAL	367	191	176	109	9	52	43	36	29	278

Auditor impact on accounting policy change

FIGURE 2.A,B

	No changes made	Some changes made	Totals
'Big 8'	97	118	215
'non-Big 8'	94	58	152
Totals	191	176	367

Incidence of policy changes and accounting firm classification

($\chi^2_1 = 9.98$, $P < .00$.)

	No changes made	Some changes made	Totals
'Big 8'	97	47	144
'non-Big 8'	94	20	114
Totals	191	67	258

Accounting firm classification and incidence of policy changes, excluding mandatory changes

($\chi^2_1 = 7.43$, $P < .00$.)

Among Big 8 firms, the Kinney classification appears to have a bearing on accounting policy changes. In Figure 2c, the numbers of policy changes are indexed against the three levels I, II, III of the Kinney classification of Big 8 accounting firms. Although a general chi-squared test is not significant ($\chi^2_2 = 1.5$), when the natural ordering of the Kinney classification categories is considered, a test based on Kendall's tau, counting numbers of concordant and discordant pairs of observations, yields $z = -1.68$ and $P < .05$. For details of the Kendall tau test see Brown (1988), and for more on the general topic of testing contingency tables with ordered categories, see Best and Rayner (1996), Beh and Davey (1999), and references therein.

The conclusion in Figure 2c is strengthened considerably if mandatory accounting changes are excluded. Figure 2d has the details. Even a general test which ignores the ordering of the Kinney classification categories yields $\chi^2_2 = 7.84$, $P < .00$, while a Kendall tau test which considers the ordered Kinney categories gives $z = 2.295$, $P = .011$.

Further analysis can be carried out on the data in Figure 2 to investigate the association of 'income changing events' with either the Big 8/non-Big 8 categorisation, or the Kinney classification of accounting firms. Because multiple 'income changing events' (i.c.es) can be associated with single firms, a different form of statistical test is required.

However, a simple analysis results if the standard Poisson model is applied to the occurrence of i.c.es. Observed cell counts are realisations of independent Poisson random variables whose parameters are products of an

FIGURE 2C

Kinney classification	No changes made	Some changes made	Totals
I (DHS, PMH, TR)	23	36	59
II (AA, AY, EW)	52	61	113
III (CL, PW)	22	21	43
Totals	97	118	215

Incidence of policy changes and Kinney classification of Big 8 accounting firms

(Kendall tau test for ordered categories contingency tables gives $z = 1.68$, $P < .05$.)

FIGURE 2D

Kinney classification	No changes made	Some changes made	Totals
I	23	15	38
II	52	30	82
III	22	2	24
Totals	97	47	144

Incidence of policy changes and Kinney classification of Big 8 accounting firms, omitting mandatory changes

(Kendall tau test for ordered categories contingency tables gives $z = 2.295$, $P = .011$.)

underlying Poisson rate with the number of firms contributing to the count. Then, using the standard fact that the distribution of a collection of Poisson variables *conditional upon their sum* is just multinomial (or binomial for just two variables), the data structure reduces to testing a single row of observed counts against an expected pattern. For this situation, a goodness-of-fit test is standard.

For example, for the 'Big 8' versus 'non-Big 8' comparison, the data in Figure 2 yields:

	Big 8	non-Big 8	Totals
Number of i.c.es	78	30	108
Number of accounting firms	215	152	367
Expected numbers	63.270	44.730	108

(Comparing 'observed' (78, 30) with 'expected' (63.27, 44.73) yields $\chi^2_1 = 8.27$, P = .00.)
For testing across the Kinney classification, the data in Figure 2 yields:

Kinney classification	I	II	III	Totals
Number of i.c.es	32	43	3	78
Number of companies	59	113	43	215
Expected numbers	21.405	40.995	15.600	78

Here, $\chi^2_2 = 15.52$, P = .0004; there is little point in applying an ordered categories test because the result is already highly significant.

Figure 3 reports substantially the same data when the auditors have been aggregated according to Kinney's (1986) classification. The distinction between Groups 1 and 2 ('structured' and 'intermediate') and Group 3 ('unstructured') are considered highly significant. Forty-two per cent of the changes allowed by Group 1 auditors are income increasing, compared to 35% of those allowed by Group 2 auditors, and only 3% by Group 3 auditors. The non-Big 8 auditors are excluded from the classification; the majority (Kinney's analysis reports 75%) would be members of Group 3, and even were they to be included as such the distinction between the extremes of the classification would remain remarkable. In addition, Figure 3 shows that Group 3 allowed only 7% of income reducing changes, compared to 10% for Group 1 firms and 11% for Group 2 firms.

The analysis of above the line and below the line changes does not take into account changes to the concept of extraordinary items since 1986; classification of changes as extraordinary items is now comparatively rare.

4. Discussion

The results clearly show that those audit firms classified as 'judgemental' in the Kinney (1986) categorisation are associated with far fewer client firms that report accounting policy choices whether these increase or decrease reported income. Within the then Big 8, around whom this investigation has been conducted, Coopers & Lybrand and Price Waterhouse appear to be less tolerant of income manipulation through accounting policy choice than their fellow auditors. However, several other factors may be contributing to the observed outcomes of this study, and they are considered here.

A number of authors (e.g., Morse and Richardson, 1983) have suggested that size of company and industrial sector will impact on the incidence of income increasing accounting policies. Eichenseher and Danos (1981) note the specialisation of auditors in particular industries. It might, therefore, be that accounting policy changes are associated with company size or industry,

FIGURE 3

KINNEY CLASSIFICATION	INCOME INCREASING		INCOME REDUCING	TOTAL CHANGES	NO. OF COMPANIES
	Above line	Below line			
GROUP 1	11	15	6	62	59
GROUP 2	23	10	10	94	113
GROUP 3	1	0	2	29	43
BIG '8'	35	25	18	185	215
NON-BIG '8'	8	11	11	93	152
TOTAL	43	36	29	278	367

Classification of income increasing/reducing changes 1986

FIGURE 4

SIZE / AUDITOR	SMALL (TA < $10m)	INTERMEDIATE	LARGE (TA > $60m)	TOTAL
Peat Marwick	11	4	3	18
Touche Ross	15	9	2	26
Deloittes	5	7	3	15
AUDITOR GP 1	31	20	8	59
Arthur Andersen	26	13	4	43
Arthur Young	15	13	10	38
Ernst & Whinney	16	14	2	32
AUDITOR GP 2	57	40	16	113
Coopers & Lybrand	10	7	8	25
Price Waterhouse	6	7	5	18
AUDITOR GP 3	16	14	13	43
TOTAL	104	74	37	215

Auditor and client size

rather than auditor. Figure 4 details the distribution of companies by size, across auditors and auditor groupings.

FIGURE 4A

Kinney Classification	Size small	intermediate	large	Totals
I	31	20	8	59
II	57	40	16	113
III	16	14	13	43
Total	104	74	37	215

Auditor and client size totals

A Kendall's tau test of association yields z = 1.546, P = .06.

FIGURE 5

AUDITOR INDUSTRY GROUPING	Group 1	Group 2	Group 3	TOTAL
Research & Consultancy	7	11	2	20
Retail & Distribution	8	13	6	27
Manufacturing & Construction	5	16	1	22
Financial & Investment	10	16	10	36
Extractive	27	45	19	91
Leisure	2	6	2	10
Non-Bank Financial Institutions	0	6	3	9
TOTAL	59	113	43	215

Distribution of auditor across industry groupings

There is some, though weak, evidence suggesting that across Big 8 auditors, an association exists between Kinney groupings and the size of client companies. While a conventional chi-squared test in Figure 4a is not significant, it can be noted that both classifications are ordinal (i.e., client size, and Kinney classification) and a Kendall tau test for association yields z = 1.546, P = .06. This P-value approaches significance, and raises the question that the Kinney classification may influence accounting policy changes indirectly through being associated with the sizes of client companies. However, this can have only a limited explanatory effect, because the strength of association throughout Figures 2a–2d is stronger than the association shown in Figure 4a.

FIGURE 6

INDUSTRY GROUP	NO. OF COMPANIES	Z-SCORE	
		MEAN	S.D.
Research & Consultancy	20	0.730	2.279
Retail & Distribution	27	0.716	2.029
Manufacturing & Construction	22	0.202	2.721
Financial & Investment	36	–0.314	2.777
Extractive	91	0.724	1.728
Leisure	10	–0.019	1.508
Non-Bank F.I.	9	1.531	1.186
TOTAL	215	0.482	2.141

Industry grouping and financial performance

FIGURE 6A

		Analysis of variance		
Source	df	SSQ	Mean SQ	F
Sectors	6	44.947	7.49	1.66 (P > 0.1)
Error	208	938.103	4.51	
Total	214	983.050		

ANOVA to test for differences in financial distress across industry groupings

The distribution of companies across the Big 8 auditors in this sample does not appear to be influenced by industry grouping of client company; the data in Figure 5, after combining the small categories *Leisure* and *Non-Bank Financial Institutions* in order to produce expected values, yields $\chi^2_{10} = 10.77$ not significant.

Cravens, Flagg and Glover (1994) suggest that firms such as Price Waterhouse and Coopers & Lybrand have a client base which is associated with lower market risk, greater profitability and lower leverage ratios. It might, therefore, be that accounting policy changes are associated with companies and industries with inferior financial performance. The z-score measures of financial distress, due to Houghton and Smith (1991), and modelled specifically for the West Australian business environment, were used to compare financial performance across auditor and industry groupings. Figure 6 details differences in mean financial performance across the seven industry groupings.

FIGURE 7

AUDITOR	NO. OF COMPANIES	Z-SCORE	
		MEAN	S.D.
Peat Marwick	18	0.784	1.590
Touche Ross	26	–0.130	1.349
Deloittes	15	0.749	1.156
AUDITOR GP 1	59	0.372	1.432
Arthur Andersen	43	–0.203	3.399
Arthur Young	38	0.018	1.746
Ernst & Whinney	32	0.790	1.621
AUDITOR GP 2	113	0.489	2.524
Coopers & Lybrand	25	0.784	1.935
Price Waterhouse	18	0.379	1.798
AUDITOR GP 3	43	0.615	1.868

Auditor and client financial performance

The Non-Bank Financial Institutions and Extractive sectors are apparently the top performers, while the Financial & Investment sector exhibits the greatest financial distress. The data in Figure 6 can be used to construct an ANOVA to test for differences in z-scores of financial distress across industry groupings (see Figure 6A).

Thus there is no evidence suggesting that financial performance differs across the industry groupings represented. Also, variation in financial performance apparently does not extend across the Kinney auditor classification. Figure 7 details differences in mean financial performance across auditor and auditor grouping.

The data in Figure 7 can be used to construct an ANOVA to test for differences in mean z-scores of financial distress across the Kinney auditor groupings (see Figure 7A).

The ANOVA in Figure 7a is not significant. However, improved financial performance is apparent as we progress from Group 1, through Group 2, to Group 3; however, the variability in Group 2 makes the intra-group differences in z-scores so large relatively that the differences between the groups are not statistically different. On an individual auditor level, mean z-scores are highest for Arthur Young (Group 2), Coopers & Lybrand (Group 3) and Deloittes (Group 1) so there is no direct correspondence between the Kinney classification of audit structure and financial performance of client.

It might be argued that the outcomes of this research lack external validity, in that they are applicable only to Western Australia, and to a period in

FIGURE 7A

Analysis of variance

Source	df	SSQ	MeanSQ	F
Between Auditor Groups	2	7.012	3.506	<1, not significant
Between Companies:				
within group 1	2	11.739	5.870	<1
within group 2	2	21.756	10.878	1.30, not significant
within group 3	1	1.717	1.717	<1
Error	206	931.487	4.522	

ANOVA to test for differences in mean levels of financial distress across the Kinney auditor classification

the late 1980s when the Big 8 still prevailed. Both issues are investigated below.

Smith (1998) re-evaluates the UK data reported by Smith (1992) to determine the link between auditor and twelve accounting manipulation techniques undertaken by the 208 largest quoted companies by market capitalisation. Smith (1998) identifies seven of these techniques to have a clear income effect, and explores the auditor connection for the 185 companies then associated with the Big 6 auditors. He notes that KPMG are associated with greater than average, and both Price Waterhouse and Coopers & Lybrand less than average, employment of pre-acquisition write-downs, and Price Waterhouse with less than average employment of extraordinary and exceptional items. Overall it is apparent that KPMG have significantly more manipulations than anticipated, and Coopers & Lybrand significantly fewer, but otherwise the direction of the auditor-effect is less clearly specified than in the findings of the present study. Certainly the 1992 UK data provides less support for the 1987 Kinney classification than the foregoing analysis.

Smith and Kestel (1999) conduct a time series analysis of accounting policy changes over the period 1988–94 for the same West Australian companies that provide the dataset for this study. However, only 49 companies survive independently across the whole period, and they make a relatively small number of policy changes (67 in all, but only 40 for the 'Big' group of auditors). The limited number of observations restrict the level of statistical analysis possible, but it is still clear that the auditor differences apparent in 1987 are not nearly so prominent across the subsequent period. The Group 3 (unstructured) auditors, Price Waterhouse and Coopers & Lybrand, had far fewer income reducing accounting policy changes than anticipated, but in other respects the three groupings are indistinguishable.

A number of studies have emphasised the importance of corporate image for the well-being of accounting firms. Scott and Van der Walt (1994) suggest that corporate image is the most important characteristic guiding firm selection

by clients; Beattie and Fearnley (1995) find that 'reputation/quality' is their most important characteristic, Armstrong and Smith (1996) that professionalism is the most important aspect of service quality to the clients of Big 6 accountants. Image is therefore an important component of accounting/auditing firms in their pursuit of diversity and product differentiation. Moizer (1998) surveyed financial directors of UK companies in both 1987 and 1996 to develop a corporate personality for the big accounting firms. He looked at a number of phrases used to describe accounting firms, and employed a semantic differential to measure the degree to which directors associated with each description. The observed diversity among firms in 1987 (much of it attributable to the extreme perceptions associated with Arthur Andersen and Deloittes, Haskins and Sells, allowed firms to be clustered into a four-group structure based on corporate image:

Group A (CL, PW, KPMG); Group B (EW, AY); Group C (DHS, TR); Group D (AA)

This grouping closely corresponds with the Kinney classification of the same year.

The corresponding 1996 survey shows Arthur Andersen still to be perceived as the 'most different' firm from its competitors, but that much of the diversity has evaporated, so that a revised clustering is more appropriate:

Group A (CL, PW, KPMG); Group B (EY, D&T); Group C (AA)

The reduction in the diversity among the world's accounting firms 1987–96 in the Moizer study mirrors the findings from the studies of accounting policy changes above. The number of major players has fallen from eight to (currently) five, and at the same time the profiles of surviving firms have come together. All of the studies cluster (CL and PW) and (EW and AY) together, suggesting a closeness of corporate cultures which might facilitate successful merger.

5. Conclusions

The results from this study identify levels of accounting income policy change associated with auditor grouping in a similar manner to that identified by Kinney (1986) and Moizer (1998) for the corresponding time period. The closeness of operations of firms within the groups on a number of activities identified may help to explain the success, or otherwise, of subsequent merger activity among the, then, Big 8.

More recent empirical evidence suggests that the degree of diversification among the largest auditing firms has declined since the late 1980s with a more structured approach now being more widely adopted. However, differences between firms and the way in which they are perceived persist.

The findings of these studies may have implications for auditor choice, auditor switching and future merger activity among auditors, and warrant further research focusing on the activities of the Big 5 worldwide.

6. References

Anderson, D. and I. Zimmer (1992a) 'Reactions to Regulation of Accounting for Goodwill', *Accounting and Finance*, Vol. 32, No. 2, pp. 27–50.

Anderson, D. and I. Zimmer (1992b) 'Time Series Analysis of Accounting Policy Choice: Reply', *Accounting and Finance*, Vol. 32, No. 2, pp. 57–60.

Armstrong, R.W. and M. Smith (1996) 'Marketing Cues and Perceptions of Service Quality in the Selection of Accounting Firms', *Journal of Customer Service in Marketing and Management*, Vol. 2, No. 2, pp. 37–60.

Beattie, V.A. and S. Fearnley (1995) 'The Importance of Audit Firm Characteristics and the Drivers of Auditor Change in UK Listed Companies', *Accounting and Business Research*, Vol. 25, No. 100, Autumn, pp. 227–239.

Beh, E.J. and P.J. Davy (1999) 'Partitioning Pearson's chi-squared Statistic for a Partially Ordered Three-way Contingency Table', *Austral NZ J Statis*, 2, pp. 233–246.

Best, D.J. and J.C.W. Rayner (1996) 'Nonparametric Analysis for Doubly Ordered Two-way Contingency Tables', *Biometrics*, 52, pp. 1153–1156.

Brown, B.M. (1988) 'Kendall's tau and Contingency Tables', *Austral J Statis*, 30, pp. 276–291.

Burns, T. and G.M. Stalker (1961) *The Management of Innovation*, Tavistock, London.

Cravens, K.S., J.C. Flagg and H.D. Glover (1994) 'A Comparison of Client Characteristics by Auditor Attributes: Implications for the Auditor Selection Process', *Managerial Auditing Journal*, Vol. 9, No. 3, pp. 27–36.

Cushing, B.E. and J.K. Loebbecke (1986) *Comparison of Audit Methodologies of Large Accounting Firms*, Studies in Accounting Research, No. 26, American Accounting Association, Savasota, Florida.

Dirsmith, M.W. and M.E. Haskins (1991) 'Inherent Risk Assessment and Audit Firm Technology: A Contrast in World Theories', *Accounting Organizations and Society*, Vol. 16, No. 1, pp. 61–90.

Dirsmith, M.W. and J.P. McAllister (1982) 'The Organic vs. The Mechanistic Audit', *Journal of Accounting, Auditing and Finance*, Fall, pp. 60–74.

Eichenseher, J. and P. Danos (1981) 'The Analysis of Industry Specific Auditor Concentration: Towards an Explanatory Model', *The Accounting Review*, July, pp. 479–492.
Fellingham, J.C. and D.P. Newman (1985) 'Strategic Considerations in Auditing', *The Accounting Review*, October, pp. 634–650.
Graham, L.E. (1985) 'Audit Risk – Part III', *CPA Journal*, October, pp. 36–43.
Joyce, E.J. and R. Libby (1982) 'Behavioral Studies of Audit Decision Making', *Journal of Accounting Literature*, Spring, pp. 103–121.
Houghton, K.A. and M. Smith (1991) 'Loan Risk and the Anticipation of Corporate Distress: West Australian Evidence', in K. Davis and I. Harper (eds), *Risk Management in Financial Institutions*, Allen & Unwin, Sydney, NSW, pp. 61–74.
Kinney, W. (1984) 'Discussants' Response to an Analysis of the Audit Framework Focusing on Inherent Risk and the Role of Statistical Sampling in Compliance Testing', in H.F. Settler and N.A. Ford (eds), *Auditing Symposium VII*, University of Kansas, pp. 127–132.
Kinney, W. (1986) 'Audit Technology and Preferences for Auditing Standards', *Journal of Accounting and Economics*, Vol. 8, No. 1, pp. 73–89.
Mian, S.L. and C.W. Smith (1990) 'Incentives for Unconsolidated Financial Reporting', *Journal of Accounting and Economics*, Vol. 12, pp. 141–171.
Mintzberg, H. (1979) *The Structuring of Organizations*, Prentice-Hall, Englewood Cliffs, NJ.
Moizer, P. (1998) 'The Corporate Images of the 1996 Big Six and the 1987 Big Eight', *Conference of the European Accounting Association*, Antwerp, April.
Morris, M. and W. Nichols (1988) 'Consistency Exceptions: Materiality Judgements and Audit Firm Structure', *The Accounting Review*, April, pp. 237–254.
Morse, D. and G. Richardson (1983) 'The LIFO/FIFO Decision', *Journal of Accounting Research*, Vol. 21, Spring, pp. 106–127.
Scott, D.R. and N.T. Van der Walt (1994) 'Choice Criteria in the Selection of International Accounting Firms', *European Journal of Marketing*, Vol. 29, No.1, pp. 27–39.
Smith, M. (1998) 'Creative Accounting: The Auditor Effect', *Managerial Auditing Journal*, Vol. 13, No. 3, pp. 155–158.
Smith, M. and J. Kestel (1999) 'A Time-Series Analysis of Accounting Policy Changes: West Australian Evidence', *School of Accounting Seminar Series*, University of South Australia.
Smith, T. (1992) *Accounting for Growth*, Century Business, London.
Stringer, K.W. (1981) 'Future Directions in Auditing Research', *The Auditor's Report*, Summer, pp. 3–4.

Sullivan, J.D. (1984) 'The Case for the Unstructured Audit Approach', in H.F. Stettler and N.A. Ford (eds), *Auditing Symposium VII*, University of Kansas.

Williams, D.D. and M.W. Dirsmith (1988) 'The Effects of Audit Technology on Auditor Efficiency: Auditing and the Timelines of Client Earnings Announcements', *Accounting Organizations and Society*, Vol. 13, No. 5, pp. 487–508.

Appendix 3: Sample Paper (2)

This paper was originally published in *Management Accounting* (UK), Vol. 77, No. 1, 1999, pp. 28–30, and is reproduced here by kind permission of the Chartered Institute of Management Accountants.

From Bean-counter to Action Hero: Changing the Image of the Accountant

Malcolm Smith and Susan Briggs
School of Accounting
University of South Australia

Abstract

Television, literature and the movies continue to perpetuate the stereotype of the boring accountant. The accounting profession does too little to change this perception so that a growing 'expectations gap' is created between what we expect of accountants and the observed reality.

The 'Monty Python' stereotype of the boring accountant was created by John Cleese (apparently in response to parents who had wished **him** to become an accountant!). The infamous 'Lion Tamer sketch' in which Cleese (as a recruitment consultant) interviews Michael Palin (accountant and would-be lion tamer) includes the following:

> An extremely dull fellow, unimaginative, timid, lacking in initiative, spineless, easily dominated. … Whereas in most professions these would be considerable drawbacks in accountancy they are a positive boon.

This line is perpetuated in both print and video, with such as 'Why Accountancy is not boring by Mr A. Putey' (Monty Python's *Big Red Book*, 1971) – a pedantic day-in-the-life exercise, focusing on cups of tea and train timetables, in a manner calculated to convey the opposite message.

Even the move from comedy to fictional novel does little to elevate the image of the accountant. Raymond Chandler (1969) describes him as:

> … a long stooped yellow faced man with bristly eyebrows and almost no chin … he had ink on his fingers and four pencils in his shirt pocket …

Alex Rogo, the management accountant in *The Goal* (Goldratt and Cox, 1984) is harassed and hen-pecked, facing the dual crises of plant closure and departing wife until the external consultant provides solutions; Bob Paget, the accountant in Fay Weldon's *Life and Loves of a She-Devil* (1983), subsequently adapted as the film *She Devil*, is the lightweight philandering husband, subjected to the scorn of a deserted wife, to the extent that she is able to frame him for tax fraud. Hardly images likely to advance the reputation of the profession and encourage the uncertain to choose an accounting career. But things get worse, especially when we look at the portrayal of the accountant in the movies.

Looking at films made before 1970 in which an accountant plays a key role we have:

1954	*Executive Suite*	Frederic March	(as Loren Shaw)
1957	*Bridge on the River Kwai*	Geoffrey Horne	(as Lt Joyce)
1959	*Room at the Top*	Laurence Harvey	(as Joe Lampton)
1959	*The Mating Game*	Tony Randall	(as Lorenzo Charlton)
1960	*The Apartment*	Jack Lemmon	(as Bud Baxter)
1968	*The Producers*	Gene Wilder	(as Leo Bloom)
1968	*The Adding Machine*	Milo O'Shea	(as Mr Zero)

The role of the accountant and the adjectives used to describe his (and pre-1970 it is always 'his') behaviour are instructive:

- Loren Shaw is a skilled businessman, but regarded as devious and calculating;
- Joyce is the hesitant lieutenant who folds under fire and whose incompetence threatens both the sabotage operation and the lives of his colleagues;
- Joe Lampton lusts after the boss' daughter, seeing marriage to her as the quickest way up the corporate ladder;
- Lorenzo Charlton (in an adaptation of H.E. Bates's *The Darling Buds of May*) is the bowler-hatted, clip-boarded IRS agent conducting a tax audit of the assets of a notorious non-taxpayer;
- Bud Baxter is the shy accounts clerk, the backroom-boy, in a New York insurance company;
- Leo Bloom is the timid accountant reviewing the books of a film product client, in search of tax loopholes;
- Mr Zero is made redundant, his accounting role taken over by an adding machine, to which he naturally responds by murdering his boss!

Devious, shy, timid and hesitant are characteristics which predominate among men described by Beard (1994) as 'lonely and dysfunctional characters'.

The post-1980 situation is different (there are at least some women accountants) but the image created is no more flattering:

1980	*Highpoint*	Richard Harris	(as Lewis Kinney)
1984	*Ghostbusters*	Rick Moranis	(as Louis Tully)
1987	*Moonstruck*	Cher	(as Loretta Castorini)
1987	*The Untouchables*	Charles Martin Smith	(as Oscar Wallace)
1988	*Midnight Run*	Charles Grodin	(as Jonathan Mardukas)
1989	*Lethal Weapon 2*	Joe Pesci	(as Leo Getz)
1989	*Look Who's Talking*	Kirstie Alley	(as Mollie)
1990	*Strike it Rich*	Robert Lindsay	(as Bertram)
1990	*Jury Duty*	Bronson Pinchot	(as Sanford)
1991	*Pure Luck*	Martin Short	(as Proctor)
1993	*Schindler's List*	Ben Kingsley	(as Itzhak Stern)
1995	*Dead Man*	Johnny Depp	(as William Blake)
1996	*Bound*	Joe Pantoliano	(as Ceasar)
1996	*Jerry Maguire*	Renee Zellweger	(as Dorothy Boyd)
1997	*For Richer or Poorer*	Wayne Knight	(as Bob Lachman)

- Lewis Kinney is an unemployed accountant, contracted to recover embezzled millions, and demonstrates super-hero skills in accomplishing the task;
- Louis Tully is the 'Keymaster', the archetypal 'nerd' accountant next door, with neither friends nor social skills;
- Loretta Castorini is a dowdy, widowed bookkeeper looking to make a sensible marriage, before overtaken by a Cinderella-like makeover;
- Oscar Wallace is the bald, vertically challenged gun-toting sidekick to Eliot Ness who audits Al Capone's ledgers for evidence of tax evasion;
- Jonathan Mardukas is accountant to the mafia, criminal and bail-jumper;
- Leo Getz demonstrates the skills the accountant can bring to the laundering of drug money;
- Kirstie Alley's 'Mollie' is an accountant having an affair with one of her clients, who falls pregnant, is dumped, and opts for single motherhood;
- Robert Lindsay's 'Bertram' is a London accountant overcome by gambling fever while on his honeymoon;
- Bronson Pinchot is yet another shy accountant, living in a sexual fantasy land, and on trial for embezzlement;
- Martin Short's 'Proctor' is described as a 'notoriously unlucky' accountant charged with the responsibility of solving the disappearance of a missing heiress;
- Itzhak Stern is the bookkeeper/cost accountant running factories for the outgoing Oskar Schindler in Nazi-occupied Poland;
- Depp's 'Blake', in a Western environment, loses his job as an accountant and reverts to gunslinging cowboy!;
- Ceasar is a money-launderer for the Mafia who is framed by two scheming lesbians;

- Dorothy Boyd is described as a 'lovely' accountant, the sole employee of fired sports-agent Tom Cruise;
- Wayne Knight (also Newman in 'Seinfeld') is an unscrupulous accountant who cooks the books and implicates his employers in a tax fraud.

There is certainly a greater representation of accountants, but dowdy, nerdish, anti-social and incompetent attributes still shine through; even the women are far from favourably portrayed. But far from 'boring', the criminal element is now to the fore with gambling, money-laundering, fraud and, generally, unprofessional behaviour well represented.

There are still no 'action heroes'; no Stallone or Schwartzeneggar in roles where the accountant saves the world, though Lewis Kinney in *Highpoint* comes close. Perhaps this is just too incredible for film producers to contemplate. Interestingly, Charles Bronson became an architect in *Death Wish* (1974), even though in Brian Garfield's novel, which formed the basis of the screenplay, he had been an accountant!

The lawyer as hero is common both from TV (*LA Law*; *The Defenders*) and film (*The Client*; *My Cousin Vinny*), so the problem is not with professionals, but with accountants.

The medical profession has had no such problem either, with successful depictions in such as *Marcus Welby*, *ER*, and *Chicago Hope*. For accountants, though, we have characters like Norm Peterson (George Wendt in *Cheers*) perennially bar-propping, and Ernie Niles (in *Twin Peaks*), the latter with the immortal line:

> I can't do that, I'm a coward, I'm a CPA.

The implications for the profession of such messages are serious. The absence of women accountants in TV and film reinforces the idea of a male-dominated profession. Even though 50% of students on undergraduate accounting courses are women, still less than 10% of the qualified accounting profession is female.

Surveys of accounting students further reinforces the stereotype. Siegel (1982) reports that such students viewed themselves as:

- honest, competent, hard-working and intelligent.

But, they did not regard themselves as:

- independent, respected, creative and boring.

However, they did attribute these latter characteristics (independence, respect, creativity and boredom) to CPAs as a professional group.

These perceptions are disturbing since, as Holt (1994) observes, if they remain uncorrected, the best and brightest students will turn to other fields rather than attempting to join the accounting profession.

Interestingly, reference to recent movie role models of accountants suggest that they are lacking in three qualities essential to success in the profession:

- communication skills;
- ethics; and
- flexibility to change.

The adoption of illegal and unethical manipulations by accountants in 1990s movies is a worrying trend, as is the continuing emphasis on white males, in subordinate positions within an organisation. Such stereotypes are powerful influences not easily overcome, especially when the profession would prefer to emphasise:

- its acceptability to allcomers;
- the predominance of accountants as CFOs;
- high levels of education;
- superior communication skills; and
- the highest ethical credentials.

There are clearly problems here, which the profession must address. The remainder of the paper considers some of the alternatives.

1. Education

Surveys of CPAs consistently reveal that employers seek the following characteristics from potential graduate employees:

- Written communication skills – the ability to present views in writing.
- Oral communication skills – the ability to present and explain.
- Effective listening skills.
- People skills – the ability to work effectively in groups.
- Conflict resolution.
- Organisational and delegation abilities.
- Creative thinking for problem-solving.
- Critical thinking and judgement.
- Ability to work under pressure of deadlines.
- Facility with the new technologies, accounting software and databases.

The list is dominated by management and interpersonal skills; accounting competence makes a fleeting appearance at number 10. The accounting

profession dictates largely the content of accounting degree courses, and as long as compliance, regulation and standards predominate the gap between 'observed' and 'required' knowledge bases will be preserved. Reassuringly, the profession is moving to place more emphasis on non-technical skills, but the tertiary education sector needs to react swiftly to counteract charges of 'irrelevance'.

2. Rebadging

Perhaps it is the words 'accounting' and 'accountant' that are causing the problem, so much so that the stereotype cannot be arrested. The diversification of the profession into 'assurance services' and 'business solutions', rather than auditing and accounting, may provide the vehicle for an appropriate refocus.

Head (1998) observes that accounting firms are positioning themselves to provide business services (not accounting) in a rapidly changing global environment. He notes that the Big 5 accountants are already replacing 'accounting' with 'professional services' and 'comprehensive consultancy' to promote their diversification into problem-solving.

3. Changing Perceptions

Edwards (1997) reports the activity of KPMG in using the movies to promote their organisation through product placements (e.g., mugs held by 'Mr Bean' and in *Wag the Dog* and airport billboards in James Bond's *Tomorrow Never Dies*).

KPMG may accept the message that perhaps society does not like accountants, but that business really needs them. (An image which lawyers have successfully cultivated.) A number of authors have suggested that it would be dangerous for accountants suddenly to be seen as flamboyant risk-takers, since this would conflict with their prudent and conservative characteristics. However, the John Harvey-Jones *Troubleshooter* TV series did much to raise public awareness of an intrusive financial consulting role (Harvey-Jones, 1992).

Certainly the profession must address the 'subordinate' issue, so common among perceptions of accountants, by emphasising the scope of their role and their primary position as strategic decision-maker. Tully (1995) suggests that 'extraordinary versatility' is what separates the super-CFO from the mere bookkeeper; he believes that a keen sense of strategy – creating value instead of just measuring it – and people skills are the essential characteristics.

Central to this argument is the need to move away from the 'bean-counting' association – the mundane recording and reporting of historical transactions – to stress the role of the accountant as manager with well-rounded strategic, marketing and interpersonal skills, heading up multinational enterprises with global interests. The generally accepted organisational domain of the accountant must be stretched. Without action there will be no heroes, and accounting runs the risk of becoming regarded as a second-class profession.

References

Beard, V. (1994) 'Popular Culture and Professional Identity: Accountants in the Movies', *Accounting Organizations and Society*, Vol. 19, No. 3, pp. 303–318.

Edwards, S. (1997) 'Here's looking at you KPMG', *CA Magazine* (Canada), June/July, pp. 14–16.

Goldratt, E.M.N. and Cox, J. (1984) *The Goal*, North River Press, New York.

Harvey-Jones, J. (1992) *Troubleshooter 2*, BBC Books, London.

Head, G. (1998) 'Is Accounting Becoming a Naughty Word?', *New Accountant*, Feb., pp. 10–16.

Holt, P.E. (1994) 'Stereotypes of the Accounting Professional as Reflected by Popular Movies, Accounting Students and Society', *New Accountant*, April, pp. 24–25.

Siegel, G. (1982) 'How Accounting Majors Perceive the Accounting Profession: A Multidimensional Scaling Approach', Midwest Meeting of the American Accounting Association, Chicago, March.

Tully, S. (1995) 'From Accountant to Action Heroes', *Time*, 27 November, pp. 64–68.

References

Abdel-Khalik, A.R. and Ajinkya, B.B. (1979) *Empirical Research in Accounting: A Methodological Viewpoint*, American Accounting Association Accounting Education Series, Vol. 4, Sarasota, FL.

Abdolmohammadi, M. and Wright, A. (1987) 'An Examination of the Effects of Experience and Task Complexity on Audit Judgments', *The Accounting Review*, January, pp. 1–13.

Abernethy, M.A., Chua, W.F., Luckett, P.F. and Selto, F.H. (1999) 'Research in Managerial Accounting: Learning from Others' Experiences', *Accounting & Finance*, Vol. 39, No. 1, pp. 1–28.

Abrahamson, E. and Amir, E. (1996) 'The Information Content of the President's Letter to Shareholders', *Journal of Business Finance and Accounting*, Vol. 28, No. 8, pp. 1157–82.

Abrahamson, E. and Park, C. (1994) 'Concealment of Negative Organizational Outcomes: An Agency Theory Perspective', *Academy of Management Journal*, Vol. 37, No. 5, pp. 1302–34.

Agnew, N.M. and Pyke, S.W. (1994) *The Science Game: An Introduction to Research in the Social Sciences*, 6th Edition, Prentice-Hall, Englewood Cliffs, NJ.

Aiken, M. and Hage, J. (1968) 'Organizational Interdependence and Intra-Organizational Structure', *American Sociological Review*, December, pp. 912–30.

Altman, E.I. (1968) 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy', *Journal of Finance*, Vol. 23, No. 4, pp. 580–609.

Altman, E.I. (1993) *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Profiting from Bankruptcy*, 2nd Edition, John Wiley, New York.

Altman, E.I., Haldeman, R. and Narayanan, P. (1977) 'Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations', *Journal of Banking and Finance*, Vol. 1, No. 1, pp. 29–54.

Altman, E.I., Marco, G. and Varetto, F. (1994) 'Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience)', *Journal of Banking and Finance*, Vol. 18, No. 3, pp. 505–29.

Anderson, M.J. (1985) 'Some Evidence on the Effect of Verbalization on Process: A Methodological Note', *Journal of Accounting Research*, Vol. 23, pp. 843–53.

Anderson, M.J. (1988) 'A Comparative Analysis of Information Search and Evaluation Behavior of Professional and Non-professional Financial Analysts', *Accounting Organizations and Society*, Vol. 13, No. 5, pp. 431–46.

Anderson, M.J. and Potter, G.S. (1998) 'On the Use of Regression and Verbal Protocol Analysis in Modeling Analysts' Behavior in an Unstructured Task Environment: A Methodological Note', *Accounting Organizations and Society*, Vol. 23, No. 5/6, pp. 435–50.

Anderson, T.W. and Rubin, H. (1949) 'Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations', *Annals of Mathematics and Statistics*, Vol. 20, pp. 46–63.

Andrews, F.M. (1984) 'Construct Validity and Error Components of Survey Methods: A Structural Modelling Approach', *Public Opinion Quarterly*, Vol. 48, pp. 409–42.

Argyris, C. (1952) *The Impact of Budgets on People*, Controllership Foundation, Inc., Cornell University, Ithaca, NY.

Armstrong, P. (1987) 'The Rise of Accounting Controls in British Capitalist Enterprises', *Accounting Organizations and Society*, Vol. 12, No. 5, pp. 415–36.

Arrington, C.E. and Francis, J.R. (1989) 'Letting the Chat Out of the Bag: Deconstruction, Privilege and Accounting Research', *Accounting Organizations and Society*, Vol. 14, No. 1/2, pp. 1–28.

Arrington, C.E. and Puxty, A.E. (1991) 'Accounting, Interests and Rationality: A Communicative Relation', *Critical Perspectives on Accounting*, Vol. 2, No. 1, pp. 31–58.

Ashton, R.H. (1974) 'An Experimental Study of Internal Control Judgments', *Journal of Accounting Research*, Spring, pp. 143–57.

Ashton, R.H. (1983) *Research in Audit Decision-Making: Rationale, Evidence and Implications*, The Canadian General Accountants Research Foundation, Vancouver.

Ashton, R.H. and Ashton, A.H. (eds) (1995) *Judgment and Decision-Making Research in Accounting and Auditing*, Cambridge University Press, New York.

Ashton, R.H. and Kramer, S.S. (1980) 'Students as Surrogates in Behavioral Accounting Research: Some Evidence', *Journal of Accounting Research*, Vol. 18, pp. 1–15.

Baiman, S. (1982) 'Agency Research in Managerial Accounting: A Survey', *Journal of Accounting Literature*, Vol. 1, pp. 154–213.

Baines, A. and Langfield-Smith, K. (2001) 'Antecedents to Management Accounting Change: A Structural Equation Approach', Biennial Management Accounting Research Conference, University of New South Wales, Sydney, February.

Baker, C.R. and Bettner, M.S. (1997) 'Interpretive and Critical Research in Accounting: A Commentary on its Absence from Mainstream Accounting Research', *Critical Perspectives on Accounting*, Vol. 18, No. 4, pp. 293–310.

Ball, R. and Brown, P. (1968) 'An Empirical Evaluation of Accounting Income Numbers', *Journal of Accounting Research*, Vol. 6, pp. 159–78.

Bandura, A. (1977) *Social Learning Theory*, Prentice-Hall, Englewood Cliffs, NJ.

Barrow, J.C. (1977) 'The Variables of Leadership: A Review and Conceptual Framework', *Academy of Management Review*, Vol. 2, No. 2 (April), pp. 231–51.

Bauman, M.P. (1996) 'A Review of Fundamental Analysis Research in Accounting', *Journal of Accounting Literature*, Vol. 15, pp. 1–33.

Bayne, C.K., Beauchamp, J.J., Kane, V.E. and McCabe, G.P. (1983) 'Assessment of Fisher and Logistic Linear and Quadratic Discriminant Models', *Computational Statistics and Data Analysis*, Vol. 1, pp. 257–73.

Beattie, V.A. and Ryan, R.J. (1989) 'Performance Indices and Related Measures of Journal Reputation in Accounting', *British Accounting Review*, Vol. 21, pp. 267–78.

Beaver, W.H. (1966) 'Financial Ratios as Predictors of Failure', *Empirical Research in Accounting*, supplement to *Journal of Accounting Research*, pp. 71–111.

Becker, S. and Green, D. (1962) 'Budgeting and Employee Behaviour', *The Journal of Business*, October, pp. 392–402.

Bedford, N.M. (1965) *Income Determination Theory: An Accounting Framework*, Addison-Wesley, London.

Bem, D.J. (1972) 'Self-Perception Theory', in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. 6, pp. 1–62.

Bennett, R. (1991) 'How is Management Research Carried Out?', in N.C. Smith and P. Dainty (eds), *The Management Research Handbook*, Routledge, London.

Berelson, B. (1952) *Content Analysis in Communication Research*, Free Press, London.

Berry, A.J., Capps, T., Cooper, D., Ferguson, P., Hopper, T. and Lowe, E.A. (1985) 'Management Control in an Area of the NCB: Rationales of accounting in a Public Enterprise', *Accounting Organizations and Society*, Vol. 10, No. 1, pp. 3–28.

Bettman, J.R. and Weitz, B.A. (1983) 'Attributions in the Board Room: Causal Reasoning in Corporate Annual Reports', *Administrative Science Quarterly*, Vol. 28, pp. 165–83.

Birnberg, J.G., Turopolec, L. and Young, S.M. (1983) 'The Organizational Context of Accounting', *Accounting Organizations and Society*, Vol. 8, No. 2/3, pp. 111–29.

Black, F. and Scholes, M. (1973) 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy*, Vol. 81, pp. 637–54.

Bloor, M. (1978) 'On the Analysis of Observational Data: A Discussion of the Worth and Uses of Inductive Techniques and Respondent Validation', *Sociology*, Vol. 12, pp. 542–54.

Blum, M. (1974) 'Failing Company Discriminant Analysis', *Journal of Accounting Research*, Spring, pp. 1–25.

Bollen, K.A. and Long, J.S. (1993) 'Introduction', in K.A. Bollen and J.S. Long (ed.), *Testing Structural Equation Models*, Sage, Newbury Park, CA, pp. 1–9.

Bonner, S.E. and Sprinkle, G.B. (2002) 'The Effects of Monetary Incentives on Effort and Task Performance: Theories, Evidence, and a Framework for Research', *Accounting Organizations and Society*, Vol. 27, No. 4/5, pp. 303–45.

Bouwman, M.J. (1984) 'Expert vs. Novice Decision Making in Accounting: A Summary', *Accounting Organizations and Society*, Vol. 9, No. 3, pp. 325–7.

Bowman, E.H. (1984) 'Content Analysis of Annual Reports for Corporate Strategy and Risk', *Interfaces*, Vol. 14, Jan.–Feb., pp. 61–71.

Boys, P.G. and Rutherford, B.A. (1984) 'The Use of Accounting Information by Institutional Investors', in B. Carsberg and M. Page (eds), *Current Cost Accounting: The Benefits and the Costs*, Vol. 2, Prentice-Hall, London, pp. 103–27.

Briers, M. and Hirst, M. (1990) 'The Role of Budgetary Information in Performance Evaluation', *The Accounting Review*, Vol. 15, No. 4, pp. 373–98.

Brin, T., Jones, M.J. and Pendlebury, M. (1996) 'UK Accountants' Perceptions of Research Journal Quality', *Accounting and Business Research*, Vol. 26, No. 3, pp. 265–78.

Broadbent, J. (1998) 'The Gendered Nature of "Accounting Logic": Pointers to an Accounting that Encompasses Multiple Values', *Critical Perspectives on Accounting*, Vol. 9, No. 3, pp. 267–97.

Brookfield, D. and Morris, R. (1992) 'The Market Impact of UK Company News Announcements', *Journal of Business Finance and Accounting*, Vol. 19, No. 4, pp. 585–602.

Brown, B.M. (1988) 'Kendall's-*tau* and Contingency Tables', *Australian Journal of Statistics*, Vol. 30, pp. 276–91.

Brown, C.E. and Solomon, I. (1993) 'An Experimental Investigation of Explanations for Outcome Effects on Appraisals of Capital Budgeting Decisions', *Contemporary Accounting Research*, Vol. 10, No. 1, pp. 83–111.

Brown, D.A., Booth, P. and Giacobbe, F. (2000) 'Organisational Influences, Ownership, and the Adoption of Activity-based Costing in Australian Firms', School of Accounting Seminar Series, University of Technology, Sydney.

Brown, L.D. (1996) 'Influential Accounting Articles, Individuals, PhD Granting Institutions and Faculties: A Citation Analysis', *Accounting Organizations and Society*, Vol. 21, No. 7/8, pp. 723–54.

Brownell, P. (1982) 'The Role of Accounting Data in Performance Evaluation, Budgetary Participation, and Organizational Effectiveness', *Journal of Accounting Research*, Spring, pp. 12–27.

Brownell, P. (1995) *Research Methods in Management Accounting*, Coopers and Lybrand, Melbourne.

Bryman, A. (2001) *Social Research Methods*, Oxford University Press, Oxford.

Buchanan, D., Boddy, D. and McCalman, J. (1988) 'Getting In, Getting On, Getting Out, and Getting Back', in A. Bryman (ed.), *Doing Research in Organizations*, Routledge, London, pp. 53–67.

Burns, T. and Stalker, G.M. (1961) *The Management of Innovation*, Tavistock, London.

Byrne, B.M. (2001) *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming*, Lawrence Erlbaum, Mahwah, NJ.

Campbell, D.T. (1965) 'Ethnocentric and Other Altruistic Motives', in D. Levine (ed.), *Symposium on Motivation*, University of Nebraska Press, Lincoln, NB, pp. 283–311.

Campbell, D.T. and Stanley, J. (1963) *Experimental and Quasi-Experimental Designs for Research*, Rand-McNally, Chicago.

Canning, J.B. (1929) *The Economics of Accountancy: A Critical Analysis of Accounting Theory*, Ronald Press, New York.

Casey, C.J. (1980) 'Variation in Accounting Information Overload: The Effect on Loan Officers' Predictions of Bankruptcy', *The Accounting Review*, Vol. 55, No. 1, pp. 36–49.

Checkland, P. (1981) *Systems Thinking, Systems Practice*, Wiley, Chichester.

Cheng, M., Luckett, P.F. and Schulz, A.K.-D. (1998) 'The Effects of Cognitive Style Diversity on Decision-Making Groups: An Empirical Analysis of a Complex Production Scheduling Task', University of New South Wales Working Paper, Sydney.

Chong, V.K. and Chong, K.M. (1997) 'Strategic Choices, Environmental Uncertainty and SBU Performance: A Note on the Intervening Role of Managerial Accounting Systems', *Accounting and Business Research*, Vol. 27, pp. 268–76.

Christensen, L.B. (1994) *Experimental Methodology*, 6th Edition, Allyn & Bacon, Boston.

Chua, W.F. (1988) 'Interpretive Sociology and Management Accounting Research – A Critical Review', *Accounting, Auditing and Accountability Journal*, Vol. 1, No. 2, pp. 59–79.

Chua, W.F. (1995) 'Experts, Networks and Inscriptions in the Fabrication of Accounting Images: A Story of the Representation of Three Public Hospitals', *Accounting Organizations and Society*, Vol. 20, No. 2/3, pp. 111–45.

Chua, W.F. (1996) 'Issues in Substantive Areas of Research: Field Research in Accounting', in A.J. Richardson (ed.), *Research Methods in Accounting: Issues and Debates*, The Canadian General Accountants Research Foundation, Vancouver, pp. 209–28.

Chua, W.F. and Degeling, P. (1993) 'Interrogating an Accounting-based Intervention on Three Axes: Instrumental, Moral and Aesthetic', *Accounting Organizations and Society*, Vol. 18, No. 4, pp. 291–318.

Clapham, S.E. and Schwenk, C.R. (1991) 'Self-serving Attributions, Managerial Cognition and Company Performance', *Strategic Management Journal*, Vol. 12, pp. 219–29.

Coffey, A. (1999) *The Ethnographic Self: Fieldwork and the Representation of Reality*, Sage, London.

Collins, F., Lowensohn, S.H., McCallum, M.H. and Newmark, R.I. (1995) 'The Relationship between Budgetary Management Style and Organizational Commitment in a Not-for-profit Organization', *Behavioral Research in Accounting*, Vol. 7, pp. 65–79.

Connole, H. (1993) *Issues and Methods in Research*, Distance Education Centre, University of South Australia, Adelaide.

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

Cooper, D.C. and Emory, C.W. (1995) *Business Research Methods*, Irwin, New York.

Cooper, D.J. and Sherer, M. (1984) 'The Value of Corporate Accounting Reports: Arguments for a Political Economy of Accounting', *Accounting Organizations and Society*, Vol. 9, No. 3/4, pp. 207–32.

Coppage, R.E. and Baxendale, S. (2001) 'A Synergistic Approach to an Accounting Educator's Primary Responsibilities', *Accounting Education: An International Journal*, Vol. 10, No. 3, pp. 239–46.

Crichton, M. (1995) *The Lost World*, Arrow Books, New York.

Cushing, B.E. and Loebekke, J.K. (1986) 'Comparison of Audit Methodologies of Large Accounting Firms', Studies in Accounting Research, No. 26, American Accounting Association, Sarasota, FL.

D'Aveni, R.A. and MacMillan, A.C. (1990) 'Crisis and the Content of Managerial Communications: A Study of the Focus of Attention of Top Managers in Surviving and Failing Firms', *Administrative Science Quarterly*, Vol. 35, pp. 634–57.

Demski, J.S. and Feltham, G.A. (1976) *Cost Determination: A Conceptual Approach*, Iowa State University Press, Ames, IA.

Dent, J. (1991) 'Accounting and Organizational Cultures: A Field Study of the Emergence of a New Organizational Reality', *Accounting Organizations and Society*, Vol. 16, No. 8, pp. 705–32.

Denzin, N.K. (1970) *The Research Act in Sociology*, Butterworth, London.

Dikolli, S. and Smith, M. (1996) 'Implementing ABC: An Australian Feasibility Study', *Australian Accounting Review*, Vol. 6, No. 2, September, pp. 45–55.

Dillman, D.A. (1978) *Mail and Telephone Surveys: The Total Design Method*, John Wiley, New York.

Dirsmith, M.W. and Haskins, M.E. (1991) 'Inherent Risk Assessment and Audit Firm Technology: A Contrast in World Theories', *Accounting Organizations and Society*, Vol. 16, No. 1, pp. 61–90.

Ditton, J. (1977) *Part-Time Crime: An Ethnography of Fiddling and Pilferage*, Macmillan, London.

Dooley, D. (1995) *Social Research Methods*, Prentice-Hall, Englewood Cliffs, NJ.

Drury, C. (1999) *Management Accounting*, 2nd Edition, McGraw-Hill, London.

Dunk, A.S. (1993) 'The Effect of Budget Emphasis and Information Assymetry on the Relation between Budgetary Participation and Slack', *The Accounting Review*, Vol. 68, No. 2, pp. 400–10.

Dunk, A.S. and Nouri, H. (1998) 'Antecedents of Budgetary Slack: A Literature Review and Synthesis', *Journal of Accounting Literature*, Vol. 17, pp. 72–96.

Edwards, R.S. (1938) 'The Nature and Measurement of Income', *The Accountant*, Vol. 99, pp. 22–6.

Eisenbeis, R.A. (1977) 'Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics', *Journal of Finance*, Vol. 22, No. 3, June, pp. 875–90.

Ezzamel, M. and Bourn, M. (1990) 'The Roles of Accounting Information Systems in an Organization Experiencing Financial Crisis', *Accounting Organizations and Society*, Vol. 15, No. 5, pp. 399–424.

Ezzamel, M. and Willmott, H. (1992) 'Accounting and Trust: Some Implications for Management Control', in T. Polesie and I.L. Johansson (eds), *Responsibility and Accounting: The Organizational Regulations of Boundary Conditions*, Studentlitteratur, Lund.

Fama, E.F. (1970) 'Efficient Capital Markets: A Review of Theory and Empirical Work', *Journal of Finance*, Vol. 25, pp. 383–417.

Ferreira, L.D. and Merchant, K.A. (1992) 'Field Research in Management Accounting and Control: A Review and Evaluation', *Accounting, Auditing and Accountability Journal*, Vol. 5, No. 4, pp. 3–34.

Ferris, K.R. (1977) 'A Test of the Expectancy Theory of Motivation in an Accounting Environment', *The Accounting Review*, July, pp. 604–15.

Festinger, L. (1957) *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford, CA.

Fogarty, T.J., Singh, J., Rhoads, G.K. and Moore, R.K. (2000) 'Antecedents and Consequences of Burnout in Accounting: Beyond the Role Stress Model', *Behavioral Research in Accounting*, Vol. 12, pp. 31–69.

Foster, G. (1986) *Financial Statement Analysis*, 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ.

Foster, G. and Young, S.M. (1997) 'Frontiers of Management Accounting Research', *Journal of Management Accounting Research*, Vol. 9, pp. 63–77.

Frazier, K.B., Ingram, R.W. and Tennyson, B.M. (1984) 'A Methodology for the Analysis of Narrative Accounting Disclosures', *Journal of Accounting Research*, Vol. 22, No. 1, pp. 318–31.

Friedman, A.O. and Lyne, S.R. (2001) 'The Beancounter Stereotype: Towards a General Model of Stereotype Generation', *Critical Perspectives on Accounting*, Vol. 12, pp. 423–51.

Friedman, M. (1953) 'The Methodology of Positive Economics', in *Essays in Positive Economics*, University of Chicago Press, Chicago.

Frydman, H.E., Altman, E.I. and Kao, D.L. (1985) 'Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress', *The Journal of Finance*, Vol. 40, No. 1, pp. 269–291.

Gallhofer, S. (1998) 'The Silences of Mainstream Feminist Accounting Research', *Critical Perspectives on Accounting*, Vol. 9, No. 3, pp. 355–75.

Gibbins, M. (1992) 'Deception: A Tricky Issue for Behavioral Research in Accounting and Auditing', *Auditing: A Journal of Practice & Theory*, Fall, pp. 113–26.

Gibbins, M. and Salterio, S. (1996) 'Experimental Accounting Research: Current Methodological Issues and Debates', in A.J. Richardson (ed.), *Research Methods in Accounting: Issues and Debates*, The Canadian General Accountants Research Foundation, Vancouver, pp. 9–24.

Gill, J. and Johnson, P. (1997) *Research Methods for Managers*, 2nd Edition, Sage, London.

Gill, J. and Johnson, P. (2002) *Research Methods for Managers*, 3rd Edition, Sage, London.

Glaser, B.G. (1992) *Basics of Grounded Theory Analysis*, Sociology Press, Mill Valley, CA.

Glaser, B.G. and Strauss, A.L. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine Publishing, Chicago.

Goldratt, E.M. and Cox, J. (1989) *The Goal*, Gower, Aldershot.

Gramling, A.A. and Stone, D.N. (2001) 'Audit Firm Industry Expertise: A Review and Synthesis of the Archival Literature', *Journal of Accounting Literature*, Vol. 20, pp. 1–29.

Gray, R. (1996) 'Some Personal Reflections on Publication, Journal "Quality" and Journal Ranking in the Academic Accounting Community', University of Dundee Discussion Paper, University of Dundee, Dundee.

Grey, C. (1996) 'On Being a Professional in a "Big Six" Firm', *Accounting Organizations and Society*, Vol. 23, No. 5/6, pp. 569–87.

Griffiths, I. (1986) *Creative Accounting: How To Make Your Profits What You Want Them To Be*, Firethorn Press, London.

Gurd, B.G. (2002) 'Grounded Theory in Accounting Research', International Graduate School of Management Working Paper, University of South Australia, Adelaide.

Hair, J.F. Jr., Anderson, R.E., Tatham, R.L. and Black, W.C. (1995) *Multivariate Data Analysis with Readings*, 4th Edition, Prentice-Hall, Upper Saddle River, NJ.

Hamer, M. (1983) 'Failure Prediction: Sensitivity of Classification Accuracy to Alternative Statistical Methods and Variable Sets', *Journal of Accounting and Public Policy*, Vol. 2, pp. 289–307.

Hamilton, D.L. and Troiler, T.K. (1986) 'Stereotypes and Stereotyping: An Overview of the Cognitive Approach', in J.F. Dovidio and S.L. Gaertner (eds), *Prejudice, Discrimination and Racism*, Academic Press, Orlando, FL, pp. 127–58.

Hammersley, M. and Atkinson, P. (1983) *Ethnography: Principles in Practice*, Tavistock, London.

Hammond, T. (1997) 'Culture and Gender in Accounting Research: Going Beyond Mynatt et al.', *Critical Perspectives on Accounting*, Vol. 8, No. 6, pp. 685–92.

Hammond, T. and Oakes, L. (1992) 'Some Feminisms and their Implications for Accounting Practice', *Accounting, Auditing and Accountability Journal*, Vol. 3, No. 3, pp. 52–70.

Harte, J.M. and Koele, P. (1995) 'A Comparison of Different Methods for the Elicitation of Attribute Weights: Structural Modeling, Process Tracing and Self Reports', *Organizational Behavior and Human Decision Processes*, Vol. 64, pp. 49–64.

Hartmann, F.G.H. and Moers, F. (1999) 'Testing Contingency Hypotheses in Budgetary Research: An Evaluation of the Use of Moderated Regression Analysis', *Accounting Organizations and Society*, Vol. 24, No. 4, pp. 291–315.

Hartmann, L. (2000) 'Ethics in Business Research: Some Issues', University of South Australia Working Paper Series, University of South Australia, Adelaide.

Harvey-Jones, J. (1992) *Troubleshooter 2*, BBC Books, London.

Harvey-Jones, J. and Massey, A. (1990) *Troubleshooter*, BBC Books, London.

Hawking, S. (1998) *A Brief History of Time*, Bantam Press, London.

Healy, P.M. (1985) 'The Effect of Bonus Schemes on Accounting Decisions', *Journal of Accounting and Economics*, April, pp. 85–107.

Heiman, V. (1990) 'Auditors' Assessments of the Likelihood of Error Explanations in Analytical Review', *The Accounting Review*, Vol. 65, No. 4, pp. 875–90.

Herzberg, F. (1966) *Work and the Nature of Man*, World Publishing Co., Cleveland, OH.

Hines, R. (1988) 'Financial Accounting: In Communicating Reality We Construct Reality', *Accounting Organizations and Society*, Vol. 13, No. 3, pp. 251–61.

Hines, R. (1992) 'Accounting: Filing the Negative Space', *Accounting Organizations and Society*, Vol. 17, No. 3/4, pp. 313–41.

Holsti, O.R. (1969) *Content Analysis for the Social Sciences and Humanities*, Addison-Wesley, Reading, MA.

Hopwood, A.G. (1983) 'On Trying to Study Accounting in the Contexts in which it Operates', *Accounting Organizations and Society*, Vol. 8, Nos 2/3, pp. 287–305.

Hopwood, A.G. (1987) 'The Archaeology of Accounting Systems', *Accounting Organizations and Society*, Vol. 12, No. 3, pp. 207–34.

Horngren, C. and Foster, G. (1991) *Cost Accounting: A Managerial Emphasis*, 7th Edition, Prentice-Hall, Englewood Cliffs, NJ.

Houghton, K.A. (1987) 'The Development of Meaning in Accounting: An Intertemporal Study', *Accounting and Finance*, Vol. 27, No. 2, November, pp. 25–40.

Houghton, K.A. (1988) 'The Measurement of Meaning in Accounting: A Critical Analysis of the Principal Evidence', *Accounting Organizations and Society*, Vol. 13, No. 3, pp. 263–80.

Houghton, K.A. and Smith, M. (1991) 'Loan Risk and the Anticipation of Corporate Distress: West Australian Evidence', in K. Davis and I. Harper (eds), *Risk Management in Financial Institutions*, Allen & Unwin, Sydney, pp. 61–74.

House, R.J. (1970) 'Scientific Investigation in Management', *Management International Review*, Vol. 4/5, No. 10, pp. 139–50.

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

Hronsky, J.J.F. and Houghton, K.A. (2001) 'The Meaning of a Defined Accounting Concept: Regulatory Changes and the Effect on Auditor Decision Making', *Accounting Organizations and Society*, Vol. 26, No. 2, pp. 123–39.

Hull, R.P. and Wright, F.B. (1990) 'Faculty Perceptions of Journal Quality: An Update', *Accounting Horizons*, Vol. 4, No. 1, pp. 77–80.

Ingram, R.W. and Frazier, K.B. (1980) 'Environmental Performance and Corporate Disclosure', *Journal of Accounting Research*, Vol. 18, No. 2, pp. 614–22.

Ingram, R.W. and Frazier, K.B. (1983) 'Narrative Disclosures in Annual Reports', *Journal of Business Research*, Vol. 11, pp. 49–60.

Ittner, C., Larcker, D.F. and Rajan, M.V. (1997) 'The Choice of Performance Measures in Annual Bonus Contracts', *The Accounting Review*, Vol. 72, pp. 231–55.

Jaworski, B.J. and Young, S.M. (1992) 'Dysfunctional Behavior and Management Control: An Empirical Study of Marketing Managers', *Accounting Organizations and Society*, Vol. 17, No. 1, pp. 17–35.

Jensen, M.C. and Meckling, W.H. (1976) 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure', *Journal of Financial Economics*, Vol. 3, pp. 305–60.

Jones, M.J. and Shoemaker, P.A. (1994) 'Accounting Narratives: A Review of Empirical Studies of Content and Readability', *Journal of Accounting Literature*, Vol. 13, pp. 142–84.

Jönsson, S. (1982) 'Budgetary Behavior in Local Government – A Case Study over 3 Years', *Accounting Organizations and Society*, Vol. 7, No. 3, pp. 287–304.

Jönsson, S. and MacIntosh, N.B. (1997) 'Cats, Rats and Ears: Making the Case for Ethnographic Accounting Research', *Accounting Organizations and Society*, Vol. 22, No. 3/4, pp. 367–86.

Jöreskog, K.G. (1969) 'A General Approach to Confirmatory Maximum Likelihood Factor Analysis', *Psychometrica*, Vol. 34, pp. 183–202.

Joyce, E.J. and Biddle, G.C. (1981) 'Anchoring and Adjustment in Probabilistic Inference in Auditing', *Journal of Accounting Research*, Vol. 19, No. 1, pp. 120–45.

Kahnemann, D. and Tversky, A. (1972) 'Subjective Probability: A Judgment of Representativeness', *Cognitive Psychology*, July, pp. 430–54.

Kaplan, R.S. (1983) 'Measuring Manufacturing Performance: A New Challenge for Managerial Accounting Research', *The Accounting Review*, Vol. 58, No. 4, pp. 686–705.

Kaplan, R.S. (1984) 'The Evolution of Management Accounting', *The Accounting Review*, Vol. 59, No. 3, pp. 390–418.

Kaplan, R.S. and Norton, D.P. (1992) 'The Balanced Scorecard – Measures that Drive Performance', *Harvard Business Review*, Jan./Feb., pp. 71–9.

Kelley, H.H. (1972) 'Attribution in Social Interaction', in E. Jones (ed.), *Attribution: Perceiving the Causes of Behavior*, General Learning Press, Morristown, NJ.

Kelly-Newton, L. (1980) 'A Sociological Investigation of the USA Mandate for Replacement Cost Disclosures', *Accounting Organizations and Society*, Vol. 5, No. 3, pp. 311–21.

Kidder, L.H. and Judd, C.L. (1986) *Research Methods in Social Relations*, Holt, Rinehart and Winston, London.

Kinney, W. (1986) 'Audit Technology and Preferences for Auditing Standards', *Journal of Accounting and Economics*, Vol. 48, No. 1, pp. 73–89.

Kirk, J. and Miller, M.L. (1986) *Reliability and Validity in Qualitative Research*, Sage, Newbury Park, CA.

Kline, R.B. (1998) *Principles and Practice of Structural Equation Modeling*, Guilford Press, New York.

Koh, H.C. (1991) 'Model Predictions and Auditor Assessments of Going Concern Status', *Accounting and Business Research*, Vol. 21, No. 84, pp. 331–8.

Kolb, D.A., Rubin, I.M. and McIntyre, J.M. (1979) *Organizational Psychology: An Experiential Approach*, Prentice Hall, London.

Kools, S., McCarthy, M., Durham, R. and Robrecht, L. (1996) 'Dimensioning Analysis: Broadening the Conception of Grounded Theory', *Qualitative Health Research*, Vol. 6, No. 3, pp. 312–30.

Krippendorff, K. (1980) *Content Analysis: An Introduction to its Methodology*, Sage, Beverley Hills, CA.

Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*, 2nd Edition, University of Chicago Press, Chicago.

Lachenbruch, P.A. (1967) 'An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis', *Biometrics*, Vol. 23, No. 4, pp. 639–45.

Lang, M.H. and Lundholm, R.J. (1996) 'Corporate Disclosure Policy and Analyst Behavior', *Accounting Review*, Vol. 71, pp. 467–92.
Langfield-Smith, K. (1997) 'Management Control Systems and Strategy: A Critical Review', *Accounting Organizations and Society*, Vol. 22, No. 2, pp. 207–32.
Larcker, D.F. and Lessig, V.P. (1983) 'An Examination of the Linear and Retrospective Process Tracing Approaches to Judgment Modeling', *The Accounting Review*, Vol. 58, pp. 58–77.
Laswell, H.D. (1948) 'The Structure and Function of Communications in Society', in L. Bryson (ed.), *The Communication of Ideas*, Harper & Row, New York.
Laughlin, R.C. (1987) 'Accounting Systems in Organizational Contexts: A Case for Critical Theory', *Accounting Organizations and Society*, Vol. 12, No. 5, pp. 479–502.
Laughlin, R.C. (1995) 'Empirical Research in Accounting: Alternative Approaches and a Case for Middle-range Thinking', *Accounting, Auditing and Accountability Journal*, Vol. 8, No. 1, pp. 63–87.
Lawler, E. (1973) *Motivation in Organizations*, Brodis/Cole, Monterey, CA.
Le Quesne, L. (1983) *The Bodyline Controversy*, Unwin, London.
LeClere, M.J. (2000) 'The Occurrence and Timing of Events: Survival Analysis Applied to the Study of Financial Distress', *Journal of Accounting Literature*, Vol. 19, pp. 158–89.
Lehman, G. (1999) 'Disclosing New Worlds: Social and Environmental Accounting', *Accounting Organizations and Society*, Vol. 24, No. 3, pp. 217–41.
Lennox, C. (1999) 'Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches', *Journal of Economics and Business*, Vol. 51, pp. 347–64.
Libby, R. (1981) *Accounting and Human Information Processing: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ.
Libby, R. and Frederick, D. (1990) 'Expertise and the Ability to Explain Audit Findings', *Journal of Accounting Research*, Vol. 28, No. 2, pp. 348–67.
Libby, R. and Lipe, M. (1992) 'Incentive Effects and the Cognitive Processes Involved in Accounting Judgments', *Journal of Accounting Research*, Vol. 30, No. 2, pp. 249–73.
Libby, R. and Luft, J. (1993) 'Determinants of Judgment Performance in Accounting Settings: Ability, Knowledge, Motivation and Environment', *Accounting Organizations and Society*, Vol. 18, No. 5, pp. 425–50.
Lillis, A.M. (1999) 'A Framework for the Analysis of Interview Data from Multiple Field Research', *Accounting & Finance*, Vol. 39, No. 1, pp. 79–105.
Littleton, A.C. (1933) *Accounting Evolution to 1900*, American Institute Publishing, New York.

Lo, A.W. (1986) 'Logit versus Discriminant Analysis', *Journal of Econometrics*, Vol. 31, pp. 151–78.
Lowe, E.A. and Shaw, R.W. (1968) 'An Analysis of Managerial Biasing: Evidence of a Company's Budgeting Process', *Journal of Management Studies*, October, pp. 304–15.
Lynn, J. and Jay, A. (1987) *Yes, Prime Minister*, BBC Books, London.
MacDonald, A.P. Jr. (1970) 'Revised Scale for Ambiguity Tolerance, Reliability and Validity', *Psychological Reports*, Vol. 26, pp. 791–8.
McClelland, D.C. (1967) *The Achieving Society*, Free Press, New York.
McConnell, D., Haslem, J.A. and Gibson, V.R. (1986) 'The President's Letter to Stockholders: A New Look', *Financial Analysts Journal*, Sept.–Oct., pp. 66–70.
MacIntosh, N.B. (1994) *Management Accounting and Control Systems: An Organizational and Behavioral Approach*, John Wiley & Sons, New York.
McLeay, S. (1986) 'Students' t and the Distribution of Financial Ratios', *Journal of Business, Finance and Accounting*, Vol. 13, No. 2, pp. 209–22.
MacKay, D.B. and Villarreal, A. (1987) 'Performance Differences in the Use of Graphic and Tabular Displays of Multivariate Data', *Decision Science*, Vol. 18, No. 4, pp. 535–46.
Mahoney, T.A., Jerdee, T.H. and Carroll, S.J. (1963) *Development of Managerial Performance: A Research Approach*, South-West Publishing, Cincinnati, OH.
Markowitz, H. (1952) 'Portfolio Selection', *The Journal of Finance*, March, pp. 77–91.
Maruyama, G.M. (1998) *Basics of Structural Equation Modeling*, Sage, Thousand Oaks, CA.
Maslow, A.H. (1954) *Motivation and Personality*, Harper & Row, New York.
Mason, J. (1994) 'Linking Qualitative and Quantitative Data Analysis', in A. Bryman and R.G. Burgess (eds), *Analysing Qualitative Data*, Routledge, London.
Mayo, E. (1933) *The Human Problems of Industrial Civilization*, Macmillan, New York.
Merchant, K.A. (1985) 'Organizational Controls and Discretionary Program Decision Making: A Field Study', *Accounting Organizations and Society*, Vol. 10, No. 1, pp. 67–85.
Merchant, K.A. and Manzoni, J.F. (1989) 'The Achievability of Budget Targets in Profit Centers: A Field Study', *The Accounting Review*, Vol. 64, No. 3, pp. 539–58.
Mia, L. (1989) 'The Impact of Participation in Budgeting and Job Difficulty on Managerial Performance and Work Motivation: A Research Note', *Accounting Organizations and Society*, Vol. 14, No. 4, pp. 347–57.

Milani, K.W. (1975) 'The Relationship of Participation in Budget Setting to Industrial Supervisor Performance and Attitudes: A Field Study', *The Accounting Review*, April, pp. 274–84.

Miles, M.B. and Huberman, A.M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook*, Sage, Thousand Oaks, CA.

Modigliani, F. and Miller, M.H. (1958) 'The Cost of Capital, Corporation Finance and the Theory of Investment', *American Economic Review*, Vol. 48, pp. 261–97.

Moizer, P. (1998) 'The Corporate Images of the 1996 Big Six and the 1987 Big Eight', Conference of the European Accounting Association, Antwerp, April.

Moses, I. (1985) 'Supervising Postgraduates', Green Guide No. 3, *Higher Education Research and Development Society of Australasia (HERDSA)*, Canberra.

Mosteller, F. and Wallace, D.L. (1963) 'Inference in an Authorship Problem', *Journal of the American Statistical Association*, Vol. 58, pp. 275–309.

Myers, C.S. (1924) *Industrial Psychology in Great Britain*, Jonathan Cape, London.

Myers, S. (1977) 'Determinants of Corporate Borrowing', *Journal of Financial Economics*, Vol. 5, No. 2, pp. 147–75.

NHMRC (National Health and Medical Research Council, Australia) (2000) 'Australian Health Ethics Committee', URL: http://www.health.gov.au/nhmrc/inside/part3c.htm.

Nunnally, J. (1978) *Psychometric Theory*, McGraw-Hill, New York.

Ohlson, J.S. (1980) 'Financial Ratios and the Probabilistic Prediction of Bankruptcy', *Journal of Accounting Research*, Vol. 18, No. 1, pp. 109–31.

Onsi, M. (1973) 'Factor Analysis of Behavioral Variables Affecting Budgetary Slack', *The Accounting Review*, Vol. 48, No. 3, pp. 535–48.

Osgood, C.E., Suci, G.J. and Tannenbaum, P.H. (1957) *The Measurement of Meaning*, University of Illinois Press, Urbana, IL.

Osgood, C.E. and Walker, E.G. (1959) 'Motivation and Language Behavior: A Content Analysis of Suicide Notes', *Journal of Abnormal and Social Psychology*, Vol. 59, pp. 58–67.

Otley, D.T. (1980) 'The Contingency Theory of Management Accounting: Achievement and Prognosis', *Accounting Organizations and Society*, Vol. 5, No. 4, pp. 413–28.

Otley, D.T. (1984) 'Management Accounting and Organization Theory: A Review of their Interrelationship', in R.W. Scapens, D.T. Otley and R.J. Lister (eds), *Management Accounting, Organisation Theory and Capital Budgeting*, Macmillan/ESRC, London.

Ouchi, W.G. (1977) 'A Conceptual Framework for the Design of Organizational Control Mechanisms', *Management Science*, Vol. 25, No. 9, pp. 833–48.

Parker, L.D. (1992) 'Questionnaires and Mail Surveys', School of Commerce Working Paper, Flinders University of South Australia, Adelaide.

Parker, L.D., Guthrie, J. and Gray, R. (1998) 'Accounting and Management Research: Passwords from the Gatekeepers', *Accounting, Auditing and Accountability Journal*, Vol. 11, No. 4, pp. 371–402.

Parker, L.D. and Roffey, B.H. (1997) 'Back to the Drawing Board: Revisiting Grounded Theory and the Everyday Accountant's Reality', *Accounting, Auditing and Accountability Journal*, Vol. 10, No. 2, pp. 212–47.

Paton, W. and Littleton, A. (1940) *An Introduction to Corporate Accounting Standards*, American Accounting Association Monograph, No. 3, New York.

Peat, Marwick & Mitchell (1976) *Research Opportunities in Auditing*, Peat, Marwick, Mitchell & Co., New York.

Pedhazur, E.J. (1982) *Regression in Behavioral Research: Explanation and Prediction*, 2nd Edition, Holt Rhinehart and Winston, New York.

Perrow, C. (1970) *Organizational Analysis: A Sociological View*, Wadsworth, Belmont, CA.

Perrow, C. (1972) *Complex Organizations: A Critical Essay*, Scott, Foreman & Co., Glenview, IL.

Peters, T.J. and Waterman, R.H. (1982) *In Search of Excellence: Lessons from America's Best-run Companies*, Harper & Row, New York.

Phillips, E.M. and Pugh, D.S. (1994) *How to Get a PhD*, 2nd Edition, Open University Press, Milton Keynes.

Popper, K.R. (1959) *The Logic of Scientific Discovery*, Hutchinson, London.

Power, M. and Laughlin, R.C. (1996) 'Habermas, Law and Accounting', *Accounting Organizations and Society*, Vol. 21, No. 5, pp. 441–65.

Power, M., Laughlin, R.C. and Cooper, D.J. (2002) 'Accounting and Critical Theory', in M. Alvesson and H. Willmott (eds), *Critical Management Studies*, Sage, London.

Preston, A. (1986) 'Interactions and Arrangements in the Process of Informing', *Accounting Organizations and Society*, Vol. 11, No. 6, pp. 521–40.

Preston, A., Chua, W.F. and Neu, D. (1997) 'The Diagnostic-related Group Prospective Payment System and the Problem of Government Rationing of Health Care to the Elderly', *Accounting Organizations and Society*, Vol. 22, No. 2, pp. 147–64.

Preston, A., Cooper, D.J. and Coombs, R.W. (1992) 'Fabricating Budgets: A Study of the Production of Management Budgeting in the National Health Service', *Accounting Organizations and Society*, Vol. 17, No. 6, pp. 561–93.

Puxty, A.G. (1993) *The Social Organizational Context of Management Accounting*, Academic Press, London.

Richardson, A.J. (1996) 'Research Methods in Accounting: Issues and Debates', Research Monograph No. 25, The Canadian General Accountants Research Foundation, Vancouver.

Robbins, S.P. (1995) *Organizational Behavior: Concepts, Controversies and Applications*, Prentice-Hall, Englewood Cliffs, NJ.

Roberts, E.S. (1999) 'In Defence of the Survey Method: An Illustration from a Study of User Information Satisfaction', *Accounting & Finance*, Vol. 39, No. 1, pp. 53–78.

Rockness, H.O. (1977) 'Expectancy Theory in a Budgetary Setting: An Experimental Examination', *The Accounting Review*, October, pp. 893–903.

Roethlisberger, F.J. and Dickson, W.J. (1939) *Management and the Worker*, Harvard University Press, Cambridge, MA.

Ronen, J. and Livingstone, J.L. (1974) 'An Expectancy Approach to the Motivational Impacts of Budgets', *The Accounting Review*, October, pp. 671–85.

Rosenhahn, D.L. (1982) 'On Being Sane in Insane Places', in M. Bulmer (ed.), *Social Research Ethics: An Examination of the Merits of Covert Participant Observation*, Macmillan, London, pp. 42–56.

Rosenthal, R. (1966) *Experimenter Effects in Behavioral Research*, Appleton Century Crofts, New York.

Ross, S.A. (1977) 'The Determination of Financial Structure – The Incentive Signaling Approach', *Bell Journal of Economics*, Vol. 8, No. 1, pp. 23–40.

Ryan, R., Scapens, R.W. and Theobald, M. (1992) *Research Method and Methodology in Finance and Accounting*, Academic Press, London.

Salancik, G.R. and Meindl, J.R. (1984) 'Corporate Attributions as Strategic Illusions of Management Control', *Administrative Science Quarterly*, Vol. 29, pp. 238–54.

Schepanski, A., Tubbs, R.M. and Grimlund, R.A. (1992) 'Within-subjects and Between-subjects Designs in Behavioral Accounting Research: An Examination of Some Issues of Concern', *Journal of Accounting Literature*, Vol. 11, pp. 121–50.

Schulz, A.K.-D. (1999) 'Experimental Research Method in a Management Accounting Context', *Accounting & Finance*, Vol. 39, No. 1, pp. 29–52.

Schumacker, R.E. and Lomax, R.G. (1996) *A Beginner's Guide to Structural Equation Modeling*, Lawrence Erlbaum Associates, Mahwah, NJ.

Scott, J. (1981) 'The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models', *Journal of Banking and Finance*, September, pp. 317–44.

Sharpe, W. (1964) 'Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk', *Journal of Finance*, Vol. 19, pp. 425–42.

Shields, J.F. and Shields, M.D. (1998) 'Antecedents of Participative Budgeting', *Accounting Organizations and Society*, Vol. 23, No. 1, pp. 49–76.

Shields, M.D. (1997) 'Research in Management Accounting by North Americans in the 1990s', *Journal of Management Accounting Research*, Vol. 9, pp. 3–62.

Shumway, T. (2001) 'Forecasting Bankruptcy More Accurately: A Simple Hazard Model', *Journal of Business*, Vol. 74, No. 1, pp. 101–23.

Sikka, P. (2001) 'Regulation of Accountancy and the Power of Capital', *Critical Perspectives on Accounting*, Vol. 12, No. 2, pp. 199–211.

Sikka, P. and Willmott, H. (1997) 'Practising Critical Accounting', *Critical Perspectives on Accounting*, Vol. 8, No. 1/2, pp. 149–65.

Silverman, D. (1985) *Qualitative Methodology and Sociology*, Gower, Aldershot.

Silverman, D. (1989) 'Telling Convincing Stories: A Plan for Cautious Positivism in Case Studies', in B. Glassner and J.D. Moreno (eds), *The Qualitative–Quantitative Distinction in Social Sciences*, Kluwer Academic Publishing, London.

Simnett, R. and Trotman, K. (1992) 'Identification of Key Financial Ratios for Going Concern Decisions', *Charter*, April, pp. 39–41.

Simon, H.A. (1959) 'Theories of Decision Making in Economics and Behavioral Science', *American Economic Review*, Vol. 49, pp. 253–83.

Simons, R. (1990) 'The Role of Management Control Systems in Creating Competitive Advantage: New Perspectives', *Accounting Organizations and Society*, Vol. 15, No. 1/2, pp. 127–43.

Slovic, P. (1969) 'Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes', *Journal of Applied Psychology*, Vol. 53, No. 4, pp. 255–63.

Smith, D. and Langfield-Smith, K. (2002) 'Structural Equation Modeling in Management Accounting Research: Critical Analysis and Opportunities', European Accounting Association Conference, Copenhagen, April.

Smith, K.J., Everly, G.S. and Johns, T.R. (1993) 'The Role of Stress Arousal in the Dynamics of the Stressor-to-Illness Process among Accountants', *Contemporary Accounting Research*, Vol. 9, Spring, pp. 432–49.

Smith, M. (1992) 'Quantification of the Trade-off between the Desirable Characteristics of Accounting Disclosures', Accounting Association of Australia and New Zealand Conference, Palmerston, NZ, July.

Smith, M. (1993) 'The Effect of Heuristics on Accounting Decision Making', *Accounting Research Journal*, Spring, pp. 38–47.

Smith, M. (1994a) 'Improving Management Accounting Reporting Practices: A Total Quality Management Approach (Part I)', *Journal of Cost Management*, Vol. 7, No. 4, Winter, pp. 50–7.

Smith, M. (1994b) 'Improving Management Accounting Reporting Practices: A Total Quality Management Approach (Part II)', *Journal of Cost Management*, Vol. 8, No. 1, Spring, pp. 49–56.
Smith, M. (1994c) 'Benchmarking in Practice: Some Australian Evidence', *Managerial Auditing Journal*, Vol. 9, No. 3, pp. 19–22.
Smith, M. (1996) 'Qualitative Characteristics in Accounting Disclosures: A Desirability Trade-off', *Managerial Auditing Journal*, Vol. 11, No. 3, pp. 11–16.
Smith, M. (1997) *Strategic Management Accounting: Issues and Cases*, 2nd Edition, Butterworth, London.
Smith, M. (1998a) 'Creative Accounting: The Auditor Effect', *Managerial Auditing Journal*, Vol. 13, No. 3, pp. 155–8.
Smith, M. (1998b) 'Conflicting Messages in Annual Reports', *Accountability and Performance*, Vol. 4, No. 2, August, pp. 43–60.
Smith, M. and Briggs, S. (1999) 'From Bean-counter to Action Hero: Changing the Image of the Accountant', *Management Accounting*, Vol. 77, No. 1, January, pp. 28–30.
Smith, M., Fiedler, B., Brown, B. and Kestel, J. (2001) 'Structure versus Judgement in the Audit Process: A Test of Kinney's Classification', *Managerial Auditing Journal*, Vol. 16, No. 1, pp. 40–9.
Smith, M. and Gurd, B.G. (2000) *Accounting Organizations and Society*, Prentice-Hall, Sydney.
Smith, M. and Kestel, J.M. (1999) 'A Time Series Analysis of Accounting Policy Changes: West Australian Evidence', School of Accounting Seminar Series, University of South Australia, Adelaide.
Smith, M. and Taffler, R.J. (1992) 'The Chairman's Statement and Financial Performance', *Accounting and Finance*, Vol. 32, No. 2, November, pp. 75–90.
Smith, M. and Taffler, R.J. (1995) 'The Incremental Effect of Narrative Accounting Information in Corporate Annual Reports', *Journal of Business Finance and Accounting*, Vol. 22, No. 8, December, pp. 1195–210.
Smith, M. and Taffler, R.J. (1996) 'Improving the Communication of Accounting Information through Cartoon Graphics', *Accounting, Auditing and Accountability Journal*, Vol. 9, No. 2, pp. 70–87.
Smith, M. and Taffler, R.J. (2000) 'The Chairman's Statement: A Content Analysis of Discretionary Narrative Disclosures', *Accounting, Auditing and Accountability Journal*, Vol. 13, No. 5, pp. 624–47.
Smith, T. (1992) *Accounting for Growth: Stripping the Camouflage from Company Accounts,* Random House, London.
Sobel, D. (1995) *Longitude*, Fourth Estate, London.
Soin, K. (1995) 'Management Control in the Financial Services Sector', in A.J. Berry, P. Broadbent and D.T. Otley (eds), *Management Control – Theories, Issues and Practices*, Macmillan, Basingstoke, pp. 283–98.

Spence, M. (1973) 'Job Market Signaling', *Quarterly Journal of Economics*, Vol. 87, No. 3, pp. 355–74.

Spicer, B.H. and Ballew, V. (1983) 'Management Accounting Systems and the Economics of Internal Organization', *Accounting Organizations and Society*, Vol. 8, March, pp. 73–96.

Staw, B.M. (1984) 'Organizational Behavior: A Review and Reformulation of the Field's Outcome Variables', in M.R. Rosenzweig and L.W. Porter (eds), *Annual Review of Psychology*, Vol. 35, pp. 627–66.

Staw, B.M., McKechnie, P.I. and Puffer, S.M. (1983) ' The Justification of Organizational Performance', *Administrative Science Quarterly*, Vol. 28, pp. 582–600.

Stock, D. and Watson, C.J. (1984) 'Human Judgment Accuracy, Multidimensional Graphics, and Humans versus Models', *Journal of Accounting Research*, Vol. 22, pp. 192–206.

Stone, P.J. and Hunt, E.B. (1963) 'Computer Approach to Content Analysis using the General Inquirer System', in E.C. Johnson (ed.), *Conference Proceedings of the American Federation of Information Processing Societies* (Baltimore, MD), AFIPS, Montvale, NJ, pp. 241–56.

Strauss, A. (1987) *Qualitative Analysis for Social Scientists*, Cambridge University Press, New York.

Strauss, A.L. and Corbin, J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage, London.

Stroebe, W. and Insko, C.A. (1989) 'Stereotype, Prejudice and Discrimination: Changing Conceptions in Theory and Research', in D. Bar-Tal, C.F. Graumann, A.W. Kruglanski and W. Stroebe (eds), *Stereotyping and Prejudice: Changing Conceptions*, Springer-Verlag, New York, pp. 3–58.

Sullivan, J.D. (1984) 'The Case for the Unstructured Audit Approach', in H.F. Stettler and N.A. Ford (eds), *Auditing Symposium VII*, The University of Kansas, Kansas City, KS.

Svenson, O. (1979) 'Process Descriptions of Decision Making', *Organisational Behaviour and Human Performance*, Vol. 23, pp. 86–112.

Swales, G.S. (1988) 'Another Look at the President's Letter to Stockholders', *Financial Analysts Journal*, March–April, pp. 71–3.

Taffler, R.J. (1983) 'The Assessment of Company Solvency and Performance Using a Statistical Model: A Comparative UK-based Study', *Accounting and Business Research*, Vol. 15, No. 52, pp. 295–308.

Taffler, R.J. (1995) 'The Use of the Z-score Approach in Practice', Working Paper 95/1, Centre for Empirical Research in Finance and Accounting, City University, London.

Tajfel, H. and Turner, J. (1985) 'The Social Identity Theory of Intergroup Behavior', in S. Worchel and W. Austin (eds), *Psychology of Intergroup Relations*, Nelson-Hall, Chicago, pp. 7–24.

Tennyson, B.M., Ingram, R.W. and Dugan, M.T. (1990) 'Assessing the Information Content of Narrative Disclosures in Explaining Bankruptcy', *Journal of Business Finance and Accounting*, Vol. 17, No. 3, pp. 390–410.

Thompson, J.D. (1969) *Organizations in Action*, McGraw-Hill, New York.

Tinker, A.M. (1980) 'Towards a Political Economy of Accounting: An Empirical Illustration of the Cambridge Controversies', *Accounting Organizations and Society*, Vol. 5, No. 1, pp. 147–60.

Tinker, A.M., Lehman, C. and Niemark, M. (1991) 'Falling Down the Hole in the Middle of the Road: Political Quietism in Corporate Social Reporting', *Accounting, Auditing and Accountability Journal*, Vol. 4, No. 2, pp. 28–54.

Tinker, A.M. and Niemark, M.K. (1987) 'The Role of Annual Reports in Gender and Class Contradictions at General Motors: 1917–1976', *Accounting Organizations and Society*, Vol. 12, No. 1, pp. 71–88.

Trotman, K.T. (1996) *Research Methods for Judgment and Decision-Making Studies in Auditing*, Coopers and Lybrand, Melbourne.

Trotman, K.T. and Wright, W.F. (1996) 'Recency Effects: Task Complexity, Decision Mode and Task-specific Experience', *Behavioral Research in Accounting*, Vol. 8, pp. 175–98.

Trubik, E. and Smith, M. (2000) 'Developing a Model of Customer Defection in the Australian Banking Industry', *Managerial Auditing Journal*, Vol. 15, No. 5, pp. 199–208.

Vroom, V.H. and Yetton, P.W. (1973) *Leadership and Decision Making*, University of Pittsburgh Press, Pittsburgh, PA.

Walker, M. and Tsalta, A. (2001) *Corporate Financial Disclosure and Analyst Forecasting Activity: Preliminary Evidence for the UK*, Certified Accountants Educational Trust, London.

Wallace, W.A. (1991) *Accounting Research Methods: Do Facts Speak for Themselves?*, Irwin, Holmwood, IL.

Watson, J.D. (1968) *The Double Helix*, Penguin Books, New York.

Watts, R.L. and Zimmerman, J.L. (1978) 'Towards a Positive Theory of the Determination of Accounting Standards', *The Accounting Review*, Vol. 53, pp. 112–33.

Watts, R.L. and Zimmerman, J.L. (1986) *Positive Accounting Theory*, Prentice-Hall, Englewood Cliffs, NJ.

Weber, R.P. (1985) *Basic Content Analysis*, Sage, Beverly Hills, CA.

Weber, R.P. (1990) *Basic Content Analysis*, 2nd Edition, Sage, Newbury Park, CA.

Weber, S.J. and Cook, T.D. (1972) 'Subject Effects in Laboratory Research: An Examination of Subject Roles, Demand Characteristics and Valid Inference', *Psychological Bulletin*, Vol. 77, No. 4, pp. 273–95.

Williamson, O.E. (1979) 'Transaction–Cost Economics: The Governance of Contractual Relations', *Journal of Law and Economics*, October, pp. 233–61.

Woodward, J. (1965) *Industrial Organization: Theory and Practice*, Oxford University Press, Oxford.

Yin, R.K. (1984) *Case Study Research, Design and Methods*, Sage, Beverly Hills, CA.

Yoon, Y. and Swales, G. (1991) 'Predicting Stock Market Performance: A Neural Network Approach', *Proceedings of the IEEE 24th Annual International Conference of Systems Sciences* (January), IEEE Computer Society Press, Hawaii, pp. 156–62.

Young, S.M. (1996) 'Survey Research in Management Accounting: A Critical Assessment', in A.J. Richardson (ed.), *Research Methods in Accounting: Issues and Debates*, The Canadian General Accountants Research Foundation, Vancouver, pp. 55–68.

Young, S.M. and Selto, F.H. (1993) 'Explaining Cross-sectional Work-group Performance Differences in a JIT Facility: A Critical Appraisal of a Field-based Study', *Journal of Management Accounting Research*, Vol. 5, pp. 300–26.

Zavgren, C.V. (1985) 'Assessing the Vulnerability of Failure of American Industrial Firms: A Logistic Analysis', *Journal of Business Finance and Accounting*, Vol. 12, No. 1, pp. 19–45.

Zmijewski, M.E. (1984) 'Methodological Issues Related to the Estimation of Financial Distress Prediction Models', *Journal of Accounting Research* (Supplement), Vol. 22, pp. 59–86.

Index