

# **Auditors' Decision Processes in Audit Planning Stage Materiality Judgments**

# Adaptive Decision Processes in Adult Learning Stage Laterality Judgments

*To Inge, Bart, Niki and Bob who were in my team when I played the game*

# **Auditors' Decision Processes in Audit Planning Stage Materiality Judgments**

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Rijksuniversiteit Limburg te  
Maastricht, op gezag van de Rector Magnificus, Prof. Dr. H. Philipsen, volgens  
het besluit van het College van Dekanen, in het openbaar te verdedigen op  
woensdag 21 december 1994 om 16.00 uur

door

**Edmond Hubert Johannes Vaassen**  
geboren te Maastricht

Promotores: Prof. Dr. J.H.R. van de Poel  
Prof. Dr. Th. J. Mock

Beoordelingscommissie: Prof. Dr. Ir. A. Kolen (voorzitter)  
Prof. Dr. W.F.J. Buijink  
Prof. Dr. D.W. Feenstra RA (Rijksuniversiteit Groningen)  
Prof. Dr. A. Schilder RA

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Vaassen, Edmond Hubert Johannes

Auditors' decision processes in audit planning stage  
materiality judgments / Edmond Hubert Johannes Vaassen. -  
Maastricht : Universitaire Pers Maastricht. - (Dissertation /  
Faculty of Economics and Business Administration,  
University of Limburg ; no. 94-22)  
Thesis Maastricht. - With ref.  
ISBN 90-5278-163-X  
Subject headings: audit judgment / human information  
processing.



# Acknowledgements

This thesis is the result of the combined efforts, ideas, advices, criticisms of many. Of course, responsibility rests entirely with me. I would like to express great appreciation to all the people who directly, or indirectly (by means of creating an environment where research might flourish) helped me in completing the study.

The team that I am indebted to, has been subject to 'personnel' changes during the years I worked on this dissertation. I will try to be as complete as possible, trying to maintain the chronological order of contribution.

The initiator of my research project (and of a lot of other projects), Jan van de Poel, encouraged me to become an auditor as well as a researcher at a very early stage of my study. Without his help, this thesis would not have been written. I would like to thank Rick Hayes (California State University) for the invaluable support he gave during the first stages of the research project. His involvement sometimes exceeded my own.

I am indebted to the public accounting firm of Deloitte & Touche in Maastricht, who allowed me to spend time doing research rather than auditing during the period I worked for them in the audit practice. A substantial financial contribution has been made by the Limperg Institute, without which data gathering and analyses would not have been possible. The type of research that is described in this dissertation necessitates the participation of test persons. Therefore, a word of appreciation goes also to the public accounting firms of Coopers & Lybrand, Deloitte Dijkster van Dien, KPMG and Moret Ernst & Young (note that data was gathered in 1988 and 1989), who allowed a number of their affiliates to participate in my research.

I would like to thank all my colleagues of the University of Limburg who provided new ideas, gave advice and criticism, and who simply created a good research environment: Jos Blommaert, Tjeu Blommaert, Laury Bollen, Henk Brink, Willem Buijink, Roger Dassen, Tom Groot, Marc Gulikers, Frank Hartmann, Harold Hassink, Gerard Mertens, Roger Meuwissen, Luc Quadackers, Ton van Reeken, Rina Vaatstra, Frans Volmer, Rolf Wenemoser, and Maurice Zinken. I am especially grateful to Bram Beek who took the time to discuss ideas and to read parts of the manuscript and to Steven Maijor for commenting in detail on an early draft of this thesis. The methodological comments of Ted Mock (University of Southern California/University of Limburg), one of my supervisors who showed me the ways to improve the manuscript, were of great value to me. I also would like to thank Caroline van Oostrom-Broers, Marcel Caubo, Rainer Steger and Frank Toebosch for the effort they put into the encoding of the transcriptions, and Stan Biggs (University of Connecticut) and Richard Baker (Fordham University, New York) for the discussions I had with them.

The editorial comments of Sen McGlinn have greatly improved the readability of

the manuscript as well as my knowledge of the English Language.

I would like to thank Professors Willem Buijink and Antoon Kolen (both University of Limburg), Arnold Schilder (University of Limburg/University of Amsterdam), and Dick Feenstra (University of Groningen) for judging the final version of the manuscript and for providing useful comments.

Finally, I would like to thank the most important members of my team: my wife Inge, for her endless optimism and our children Bart, Niki and Bob just for being there.

Eddy Vaassen  
Maastricht, October 1994

# Contents

## 1. Introduction

1.1	The development of the modern auditing function	11
1.1.1	History	12
1.1.2	Risk-sharing	12
1.1.3	Approaches to audit research	13
1.2	The audit process	14
1.2.1	Audit risk models and planning stage risk assessment	14
1.2.2	Audit planning stage materiality assessment	16
1.2.3	Planning stage materiality and risk assessment: a combined view	16
1.2.4	The nature of audit decision tasks and planning stage materiality and risk assessment	19
1.3	Research questions	20
1.4	A general audit judgment model	22
1.5	Purpose and outline of the study	25
1.5.1	Issues in audit judgment research	25
1.5.2	Outline of the study	26
	Notes	27

## 2. An overview of audit judgment research

2.1	Introduction	29
2.2	Professional judgment in auditing	29
2.3	A framework for audit judgment literature	31
2.4	Studies of pre-decisional behavior	32
2.5	Expert system development	34
2.6	Audit standard setting and regulation at the firm level	35
2.7	Personality and cognitive style	37
2.8	Lens model studies	37
2.9	Probabilistic judgment models and the subjective expected utility paradigm	38
2.10	Positioning the current research relative to previous research	40
	Notes	42

<b>3.</b>	<b>Technical aspects of personality assessment, verbal protocol analysis, and expert system development in relation to audit judgment</b>	
3.1	Introduction	45
3.2	Personality and cognition	45
3.2.1	Auditors' accounting and auditing knowledge	47
3.2.2	Personality tests	49
3.3	Capturing human judgment in verbal protocols	54
3.3.1	Methods for VPA	55
	<i>Concurrent verbalization</i>	55
	<i>Retrospective verbalization</i>	55
3.3.2	Levels of verbalization	58
3.3.3	Techniques in VPA	59
	<i>Objective elicitation of cognitive structures</i>	60
	<i>Uniform coding systems</i>	61
3.4	Expert modelling as a method for studying audit judgments	64
3.4.1	Determination of expertise	64
3.4.2	Knowledge elicitation techniques	66
	<i>Interviews</i>	66
	<i>Observation</i>	68
	<i>Multidimensional techniques</i>	68
	<i>Verbal protocol analysis and its application in expert system development</i>	69
	<i>Knowledge representation</i>	70
3.5	Synthesis	70
	Notes	72
<b>4.</b>	<b>Research methodology</b>	
4.1	Introduction	75
4.2	Research design	75
4.2.1	Subjects	75
4.2.2	Procedures	76
4.3	Task design	77
4.4	Measurement of information access sequence and attention duration	81
4.4.1	Information access sequence	81
4.4.2	Information attention duration	83
4.5	Verbal protocols as data	83
4.5.1	The development of a technique for transforming verbal protocols into data	83
4.5.2	Relevant statement, operator and task encodings	91
4.6	Research questions	92
4.6.1	Psychological profiles and demographic characteristics	92

4.6.2	Psychological profiles, demographic characteristics and problem-solving behavior	98
4.6.3	Auditors' materiality knowledge	103
	Notes	104
<b>5.</b>	<b>Auditors' psychological profiles in relation to demographic characteristics</b>	
5.1	Introduction	105
5.2	Results	105
5.2.1	Psychological profiles	105
5.2.2	The relationships between demographics and psychological profiles	107
5.3	Summary and conclusions	110
	Notes	113
<b>6.</b>	<b>Auditors' psychological profiles, demographic characteristics, and problem-solving behavior</b>	
6.1	Introduction	115
6.2	Results	115
6.2.1	Operators and task encodings	115
6.2.2	Recall and completeness of information access	121
6.2.3	Agreement on information access sequence and attention duration	129
6.3	Summary and conclusions	136
	Notes	141
<b>7.</b>	<b>Auditors' knowledge representation at the knowledge level</b>	
7.1	Introduction	143
7.2	Knowledge representation	144
7.3	Results	145
7.4	Summary and conclusions	148
	Notes	151
<b>8.</b>	<b>Summary and discussion</b>	
8.1	Summary	153
8.1.1	Summary of research methodology	153
8.1.2	Summary of results	154
	<i>Results concerning psychological profiles (Q1)</i>	154
	<i>Results concerning psychological profiles in relation to</i>	

	<i>problem-solving behavior (Q2)</i>	154
	<i>Results concerning materiality determining factors (Q3)</i>	156
8.2	Discussion	156
8.2.1	Practical implications	156
8.2.2	Theoretical implications	158
8.2.3	Limitations of the research	159
8.2.4	Suggestions for further research	160
8.3	Concluding remarks	164

## Appendices

Appendix A	The AICPA audit risk model and the Leslie/ CICA audit risk model	166
Appendix B	The mathematical relationship between materiality and risk	168
Appendix C	Audit Judgment Research	170
Appendix D	Historical context of verbal protocol analysis	181
Appendix E	Myers-Briggs Type Indicator Questionnaire	186
Appendix F	Tolerance For Ambiguity Questionnaire	188
Appendix G	Research questions and their interrelationships	190
Appendix H	VPA Case description Midcom Ltd	191
Appendix I	Information page standard ranking, average ranking, and attention duration	193
Appendix J	Protocol encoding	194
Appendix K1	Task structure materiality threshold determination	196
Appendix K2	Semantic network materiality threshold determination	198
Appendix K3	Method in materiality threshold determination	199
Appendix L	Structured cases used for system examination	201

<b>Bibliography</b>	205
---------------------	-----

<b>Nederlandse Samenvatting (Summary in Dutch)</b>	221
--	-----

<b>Curriculum Vitae</b>	227
-------------------------	-----

## Introduction

---

### 1.1 The development of the modern auditing function

Auditing is the process by which a competent, independent person (the auditor) accumulates and evaluates evidence about quantifiable information related to a specific economic entity (e.g. a firm) for the purpose of determining and reporting on the degree of correspondence between the quantifiable information and established criteria [Arens & Loebbecke, 1991].

When considering this widely accepted definition, two important questions emerge. The first is *why* audits are performed and the second is *what* main task elements constitute an audit. These questions implicitly give rise to two other questions: *what* conditions should be met to perform audit tasks effectively and efficiently, and *how* does the auditor acquire and process audit evidence. *These two questions, one regarding the conditions for improved audit decision quality and the other regarding the decision processes that eventually culminate in audit decisions, are the subject of this dissertation.* Questions as to *why* audits are performed and *what* task elements constitute an audit will be discussed briefly in this introductory chapter in order to put the research topics of this dissertation in their proper, larger context, and in order to explain why specific research questions will be developed.

The question of why audits are performed can conveniently be answered from a historical point of view. The evolution of the Latin term 'audit', from the ancient interpretation of 'he hears' to the dominant, modern interpretation of audits as tests of financial statements, performed by independent certified public accountants (in the Netherlands: *Registeraccountants*), will then be discussed briefly in section 1.1.1.

The question of which task elements constitute an audit can best be answered in the course of analyzing the contractual relationships between the client, the stakeholders, and the auditor, applying concepts from agency theory such as 'risk sharing' and 'principal agent relationships' (see Watts & Zimmerman [1983] and Chow, Kramer & Wallace [1988]). The fact that auditors have to deal with risks in their relationships with clients and stakeholders is relevant in this study, because it indicates that auditors must have the skills and knowledge relevant to risk assessment tasks.

When assessing risk, an assessment must be made regarding the object of risk. For example, consider the statement: 'there is a great risk that there is an error which exceeds one thousand guilders'. In this statement, the object of risk is 'an error which exceeds one thousand guilders'. In audit terminology, this amount is referred to as the materiality threshold. Section 1.2.3 provides a more elaborate discussion of the interrelationship between risk and materiality. The auditor must also have skills and knowledge relevant to materiality assessment tasks.

### **1.1.1 History**

In ancient societies it was common practice for the accounts of an estate, domain or manor to be read aloud to those in authority (e.g. the landowners) by those who had compiled them [see e.g. Woolf, 1986, p.1]. The compilers were accompanied by other persons, who also knew the contents of the accounts to be read aloud and who verified whether or not those messages were correctly read aloud. This is the closest interpretation of the Latin word 'audit'. The role of the auditor was then played by the accompanying person. In this setting, the auditor would be known for his integrity and independence of mind. If the auditor approved the accounts, he gave them extra credibility.

Two important milestones in the history of auditing are worth mentioning here. Firstly, the rise of capitalism in the place of ecclesiastical and royal hierarchies and the guilds. Capitalism emphasizes exchange relationships rather than social networks. Secondly, the separation of management and ownership initiated at the end of the 16th century in the Low Countries, and which became a widespread phenomenon in the industrial revolution in the 19th century. These two developments explain the emergence of problems of trust, reliability, protection against self-seeking behavior, cheating, etc. (see Jenssen & Meckling [1976], Watts & Zimmerman [1983] and Van de Poel [1988] for more elaborate reflections on these problems and their historical relevance). However it was not until the 19th century, when financial accounting became more and more complex, that auditing ceased to be performed by laymen. Even where there was no question about the independence and integrity of the compiler and auditor, the shareholders or other owners were no longer able to understand what was being called out to them. So the need emerged for a specialized function: the auditing function as we know it today.

### **1.1.2 Risk-sharing**

The external effect of auditing is to reduce the uncertainties that outside stakeholders, notably investors, are exposed to because of possible misrepresentations, distortions, or strategic or opportunistic information biases on the part of the managers who have to report to them. If they are to reduce



uncertainties efficiently, auditors may have to share some of the risks with the other two parties, in terms of indemnities for investors against any false reporting by management. Therefore, auditors need to possess specific skills and knowledge with respect to:

- risk assessment and risk control, that is, they need to know how to detect risks and misrepresentations and how to correct them;
- financial accounting techniques and regulations.

Several institutional arrangements exist to reduce competition and collectively reduce audit risk, such as professional bodies (NIVRA, AICPA, ICAEW, etc.), codes of conduct, and schemes for training and evaluating new members. Auditors are also involved in the regulation of financial accounting, thus reducing risks on a preventative basis (and creating commercial opportunities). All these arrangements and governance structures make the audit market less than fully competitive [Antle, 1982; Watts & Zimmerman, 1983]. Yet, even in an imperfectly competitive market, auditors need to have skills and knowledge with respect to price-setting, taking into account the ultimate professional risk they run, of having to meet compensation claims against them. The role of professional risk in auditing should never be underestimated. This dissertation is therefore based on the proposition that auditors need to have expertise in risk and materiality assessment, in order to maintain their position in a dynamic and competitive market.

### 1.1.3 Approaches to audit research

Audit research plays an important role in the competitive market for audit services. Audit efficiency and effectiveness have clearly benefited from research findings in recent decades. (e.g. the development of the audit risk model [AICPA, 1972], and the formalization of audit approaches in manuals of public accounting firms [Cushing & Loebbecke, 1986]). It is conventionally sub-divided into audit *market* research and audit *process* research. Audit market research investigates how supply and demand meet in the market for audit services, and includes the history and role of auditing and public accounting firms, while audit process research investigates how audit services are supplied by auditors. This specialization is quite pragmatic, as is demonstrated by Abdel-khalik & Solomon [1988]. Strictly speaking, this dissertation is about audit process research. However, when studying audit processes, the context in which these processes take place cannot be neglected, lest the research become fragmentary and so have no practical relevance.

In his inaugural lecture, Van de Poel [1988] drew attention to two other strands in audit research:

*'We are dealing with ... two research approaches, or theories if you like. The first one is cognitive accounting research, the other is 'agency theory'. In the former,*

*the emphasis is on potential misinterpretations of signals, in the latter the emphasis is on intentional or unintentional misrepresentations aimed at personal gain. Both approaches are relevant for the analysis of accounting processes, but they are seldom considered simultaneously.'*

A cross-classification of both the audit process and audit market distinction and the cognitive accounting and agency theory distinction would reveal a close relationship between audit process and cognitive accounting research and between audit market and agency theory research. So, cognitive accounting research and agency theory research are as related as audit process and audit market research.

## 1.2 The audit process

The audit process is generally conceived of as consisting of three main stages: planning, fieldwork, and the evaluation of audit evidence (see Cushing & Loebbecke [1986], Wallage [1991]). Evidently, decisions in the planning stage have an impact on the subsequent fieldwork and the audit evidence to be evaluated [Joyce, 1976]. These planning stage decisions involve mainly *materiality assessments*, *risk assessments* and the resulting *evidential planning* decisions (nature, extent, and timing of audit procedures). There is widespread recognition within the auditing profession that evidential planning decisions are critical to the effectiveness and efficiency of an engagement [Mock & Wright, 1992]. This dissertation focuses on these decisions, in particular on materiality and risk assessment in the planning stage of financial statement audits.

### 1.2.1 Audit risk models and planning stage risk assessment

Prior probability assessment in auditing is concerned with determining the magnitude of several risks that the auditor faces when accepting an audit engagement. Two audit risk models have been proposed, applied and criticized in recent decades: first, the joint risk model of the AICPA (see for example AICPA [1988]):

$$AR = IR * CR * DR$$

and second, the Leslie/CICA model [1985]:

$$AR = \frac{IR * CR * DR}{IR * CR * DR + (1 - IR)}$$

Where:

AR = audit risk

IR = inherent risk

*CR* = internal control risk

*DR* = detection risk

The differences between the two models are analyzed in detail in appendix A.

These two audit risk models evidently consist of the same components, *AR*, *IR*, *CR*, and *DR*, but combined in different ways. The auditor has to estimate the magnitude of these components. However, how sure can he be that his risk assessment is correct? In other words, how reliable is the information that leads the auditor to set probabilities at a certain level? Unreliability may be affected by factors such as: the reliability and the credibility of the source, the relevance of the information, or the presence of conflicting evidence [Joyce & Biddle, 1981b; Bamber, 1980, 1983; Mock & Washington, 1989; Bedard, 1990, 1991a; Moeckel, 1991]. The quality of auditors' risk assessments depends of course on their ability to identify unreliable information sources, irrelevant information, and contradictory evidence.

The importance of a proper risk assessment becomes even clearer when investigating the meaning of the different risk categories and the consequences of a specific risk assessment.<sup>1</sup> Audit risk is the chance that the auditor may issue an unjustified unqualified opinion. Normally this is an input variable because it is *a priori* fixed at an acceptable level (e.g. 5%). So a better term would be 'maximum tolerable audit risk'. Inherent risk is the likelihood that an error exists prior to the corrective influence of internal controls. It is because of the existence of this risk that an internal control system is set up and an auditor is engaged. Internal control risk is the likelihood of the internal control system not detecting material errors. Detection risk is the chance that the auditor will not detect material errors. If the maximum tolerable audit risk is set at a low level, meaning that the auditor is only willing to accept a low audit risk, this implies that the other elements constituting the audit risk model must compensate for this to achieve an acceptable level of efficiency. Inherent risk cannot be altered by the auditor because it is determined by client-specific circumstances. Internal control risk can only be altered in the long run, by the implementation of the auditor's internal control advice. Detection risk is the only element of any audit risk model that can be influenced in the short run (i.e. during the audit process) by the auditor. Since an audit program consists of a diverse set of audit procedures, detection risk can be subdivided into a set of distinguishable risks. One well known division is into *analytical review risk* versus the risk entailed in making *substantive tests of details*. Another classification, usually applied in audit sampling, is *sampling risk* versus *non-sampling risk*. An important element that is often omitted in defining the sub-categories of detection risk is the risk that remains because the auditor can make errors during the audit process, in other words, the risk that remains because the auditor is human.

### 1.2.2 Audit planning stage materiality assessment

In auditing, a material error is one that, by itself or together with other errors, would have an impact on decisions taken by a rational reader of financial statements [Leslie, 1985]. So the auditor finds himself in the position of the rational reader. He must determine what kinds of decisions are to be taken on the basis of the financial statements, so as to determine what amount would have an impact on those decisions. However, it is rather doubtful whether a rational reader of financial statements is able to determine the materiality level as defined here. Presumably the rational reader does not think in terms of materiality in the way auditors do. Similarly, he does not think in terms of confidence intervals when reading balance sheet and profit and loss items, in the way auditors do. Once an auditor's opinion is given on financial statements, the amounts stated are considered to give a true and fair view of the underlying business process. Here we arrive at a paradoxical situation: the auditor must think as if he were the rational reader of financial statements but the readers would not normally be considered rational in the auditor's sense. The auditor would have trouble in finding any 'rational reader' or determining what decisions such a reader would take on the basis of the audited financial statements.

In response to this paradox, both academics and professional auditors have put tremendous effort into the search for an objective method of determining materiality thresholds. As could be expected, the true materiality level has not been found, but, as the next best alternative, a set of quantitative guidelines has been developed. These guidelines have been issued on all levels, from audit firms to international standard-setting bodies [Leslie, 1985]. However, despite the mechanistic and quantitative elements in these guidelines, the individual auditor has always had the option of adapting the calculated materiality level to client-specific circumstances. This allows room for the use of heuristics when determining materiality in auditing.

### 1.2.3 Planning stage materiality and risk assessment: a combined view

Materiality is thus not one concept which is uniformly applicable all over the world for all kinds of businesses. More than that, it is a relative concept [Leslie, 1985] from two points of view: first, it can be expected to be dependent on the size of the business and second, it can be expected to be dependent on the auditor's perception of client-specific problem areas. Some of these client-specific problem areas are part of the inherent risk: the more extensive the client-specific problem areas, the higher the inherent risk and the more audit effort is needed (*ceteris paribus*). However, some other client-specific problem areas are true materiality determining factors: the higher the materiality threshold, the less audit effort is needed (*ceteris paribus*). This means that materiality and risk both contribute to determining the nature, extent and timing of audit procedures. However, there is one theoretical

objection to considering risk and materiality as substitutable concepts in forming an auditor's opinion. In statistical terms, materiality deals with the concept of accuracy whereas risk deals with the concept of confidence (more precisely: the complement of confidence). If for example the auditor decides to perform the audit accepting a maximum 5% audit risk that the maximum error is 10%, then the risk of the hypothesis, 'The financial statements do not contain an error exceeding 10%' being incorrectly accepted is 5%. In statistical terms, the beta-risk is 5%.

The relationship between accuracy and confidence in auditing can be stated as follows: the more accurate the audit is required to be, under unchanged confidence requirements (i.e. maximum tolerable detection risk) the more audit effort should be undertaken, and, the more confident the auditor is required to be that the audited financial statements do not contain any material error, under unchanged accuracy requirements, the more audit effort should be undertaken. This relationship is represented graphically in figure 1. This figure is based on sampling theory. In appendix B the mathematical relationship between materiality and maximum tolerable detection risk<sup>2</sup> is given.

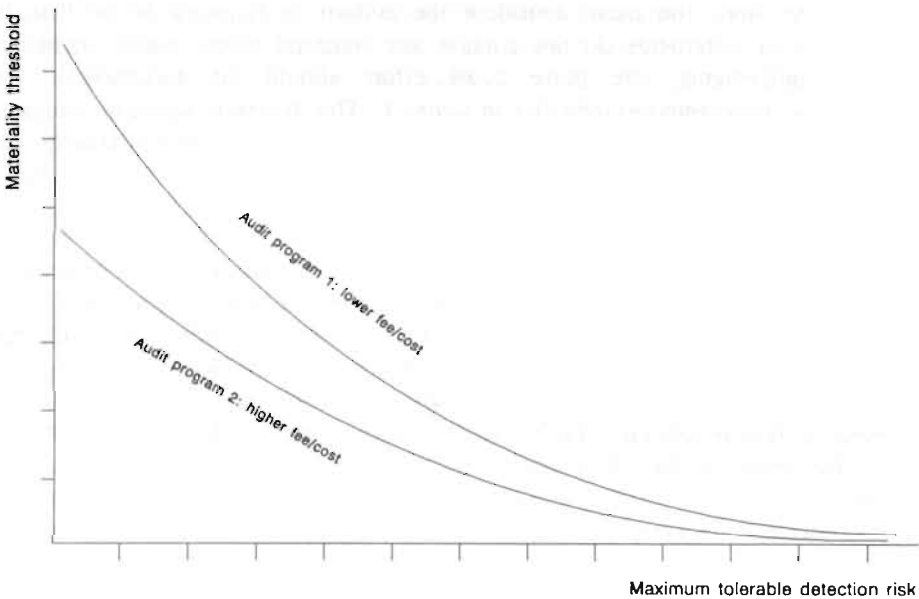
Each of the curves in the figure represents a set of materiality/risk combinations that can be achieved with identical audit-programs. The nature, extent and timing of audit procedures as represented by the upper curve is such that a lower fee is calculated, as compared with the lower curve. However, there is one practical objection to quantifying the relationship between materiality, audit procedures and detection risk in this way: with the exception of an application in audit sampling (the determination of sample sizes in tests of balance details), hardly any empirical research has been conducted in this field. In particular, the relationship between materiality and detection risk on the one hand, and the nature, extent and timing of audit procedures (other than tests of details of balances) on the other hand has only partially been assessed [Mock & Wright, 1992], and then in non-empirical deductive studies. In view of the analysis above, it is clear that planning stage materiality and risk both have an impact on the nature, extent and timing of audit procedures. When starting an audit, the auditor faces financial statements that may contain material errors. The prior probability of these statements containing material errors can be split into an inherent risk and an internal control risk. Regardless of the risk model that is used, a higher prior risk results in more audit effort. More audit effort can be induced by:

1. a reduction in the maximum tolerable detection risk (because this is a function of prior risk),
2. a reduction in the materiality threshold.

So although risk and materiality relate to different concepts (respectively confidence and accuracy), their impact on the nature, extent and timing of audit procedures is similar.

It is easy to show, by means of an example, that materiality and risk are determinants of the efficiency and effectiveness of an audit. Suppose that a loan officer is asked to approve a loan to a firm. He wants the auditor of that firm to issue an opinion on the financial statements in order to reduce the risk that the financial information on the basis of which he will make the decision is inaccurate. The auditor then investigates what risks must be recognized regarding the existence of material misrepresentations in the financial statements. The larger the risks, the more audit effort is needed to mitigate these risks. However, the question of what

**Figure 1:** The relationship between materiality and maximum tolerable detection risk.



amount, if misrepresented, must be corrected (and hence must first be detected) is as important as is the risk assessment, because the smaller the misrepresentations are which must be corrected, the more audit effort is needed in order to find that (greater number of) misrepresentations. So, the lower the materiality threshold, the more audit effort is put into the engagement. In this specific situation, the height of the materiality threshold is entirely dependent upon the amount the loan officer is willing to accept. However, audits are usually not conducted at the request of one stakeholder, but rather on the basis of the unspoken, generic request of the stakeholders as a whole. This makes materiality assessment a difficult exercise, because meeting the needs of all stakeholders is usually a matter of compromise.

As has been demonstrated, regardless of the risk model that is being used, there is

a trade-off between risk and materiality thresholds. So materiality and audit risk need to be considered together in the planning stage of an audit, a logical implication which has been recognized by regulatory bodies [AICPA, 1988]. In this dissertation, auditor problem-solving behavior is studied while the subjects are performing a combined materiality and risk assessment task.

#### **1.2.4 The nature of audit decision tasks and planning stage materiality and risk assessments**

Sniezak & Henry [1989] define three broad categories of decision tasks based on the degree of uncertainty about the accuracy of outcomes. If decision-makers can be completely certain about the accuracy of their decisions, then the task is labelled a *knowledge task*. If decision-makers remain completely uncertain about the accuracy of their decisions, then the task is labelled a *guess*. Finally, if some uncertainty remains, and the degree to which decision-makers have confidence in their decisions significantly determines 'if' and 'how' that decision will be used, then the decision task is labelled a *judgment*.

Most auditing tasks are judgments, because auditors can hardly ever be certain that they made the right decision. This also applies to the planning stage materiality and risk assessments that will be studied in this dissertation. What amount should be considered material? The auditor can only determine that amount by trial and error. Unfortunately, in auditing practise there are no second chances. Once the auditor has issued an opinion on audited financial statements, he cannot change it in the same fiscal year's statements. Therefore, if stakeholders require a more accurate audit on the basis of the financial statements after audit - which are the only financial statements which are publicly available - the auditor cannot comply without creating problems for himself.

Auditor problem-solving behavior, as a research field, is considered promising for a number of reasons: firstly, since professional judgment is seen as an integral (and presumably the most important) part of auditing, the underlying decision processes are worth studying. Secondly, audit practitioners should be keenly interested in the research findings because high standards of field performance require continuous education and search for better methods. Finally, audit judgment, although it is a complex problem area, can easily be subdivided into constituent fields which are less complex for researchers to study.

The remainder of this chapter first describes the research questions addressed in this dissertation and their rationale (§1.3). Next, it introduces and explains a general audit judgment model (§1.4). Finally, the purpose and outline of the study is given (§1.5).

### 1.3 Research questions

Understanding actual decision-making is a prerequisite for improving decisions [van de Poel, 1986]. Waller & Jiambalvo [1984] present arguments for the use of normative models in human information processing research (HIP research) as well as the counter-arguments in favor of descriptive research in this field. Waller & Felix [1984] suggest that a program of research into how an auditor learns from experience could proceed as follows: (1) descriptive research, which examines how and how well the auditor acquires, represents or uses experience-based knowledge, and (2) prescriptive research, which includes the construction of instructional or decision aids which might improve the processes by which the auditor acquires and stores experiential data.

Generally, expertise is considered to consist of formal education, on-the-job experience, and innate abilities [Bonner & Lewis, 1990]. There is no doubt that real experts outperform non-experts. Formal education and on-the-job experience can easily be measured, but the innate abilities which make an 'expert' out of an experienced, well-trained person are more difficult to measure. An explorative approach to subjects' problem solving behavior, which links personal differences to observed differences in problem-solving behavior, then becomes interesting in order to find the determinants of expertise and hence of decision quality.

Based on the thoughts voiced by Waller & Jiambalvo [1984], Waller & Felix [1984], and Van de Poel [1986], three questions relating to auditors' judgments in making planning stage materiality and risk decisions are addressed in this dissertation.

#### Research question 1:

Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?

Auditors' psychological profiles can be measured by means of self-assessment questionnaires or by means of simple pencil and paper tests. These questionnaires and tests will be discussed in chapter 3. In this study, an auditor's psychological profile is considered to consist of scores on the dimensions of information acquisition style, information processing style, and tolerance for ambiguity. The resulting psychological profiles are related to the number of years of audit experience, firm affiliation, and educational background. In this dissertation, audit experience, firm affiliation, and educational background together constitute a set of demographic characteristics.

We first drew up an overview of auditors' psychological profiles and demographic



characteristics, in order to determine what psychological profiles are prevalent within the population of auditors in the project, and then studied how these psychological profiles are related to the demographic characteristics just mentioned. The latter step is taken because, if it is possible to identify covariability between psychological and demographic variables, it might be possible to simplify any audit judgment model to be developed.

#### Research question 2:

Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?

When investigating this research question in relation to existing audit judgment literature [e.g. Choo, 1989; Colbert, 1989; Davis & Solomon, 1989; Pincus, 1991], it appears that there are three main themes that need further attention: audit expertise, auditor consensus, and auditors' psychological profiles.

#### *Audit expertise*

On the basis of earlier research findings regarding the development of episodic memory [e.g. Kolodner, 1983; Anzai, 1987; Einhorn, 1974] we can presume that novices will use less heuristics than experts and hence will need more information and intermediate thought processes to arrive at a solution to a particular problem. This proposition is tested in this study while trying to answer the question regarding what factors constitute audit planning stage materiality and risk expertise.

#### *Auditor consensus*

Wright [1988] lists more than 30 studies which treat audit expertise from different points of view. Audit practice is concerned with problems of competition: how to audit as efficiently and effectively as possible in order to be able to charge lower fees and to provide superior quality. So, audit expertise is directly linked to audit quality. Here the problem of an objective determination of the nature of audit quality emerges. Since most audit decisions are to be classified as judgments [Pincus, 1991], no objective criterion for measuring audit decision quality is available. The consensus criterion as a substitute measure for decision quality has been proposed by Einhorn [1974] and has been applied extensively since [e.g. Joyce, 1976; Moriarity & Barron, 1979; Ashton & Kramer, 1980; Ashton & Brown, 1980; Mock & Turner, 1979, 1981; Gaumnitz *et al.*, 1982; Hamilton & Wright, 1982; Krogstad *et al.*, 1984; Bedard, 1991b]. Ashton [1985] was the first to investigate the empirical relationship between consensus and accuracy. It was found that the correlation between the two attributes was strongly positive, and hence consensus might be an adequate substitute measure when decision-making is highly judgmental. Furthermore, it seems that various efforts (audit manuals,

continuing education, uniform exams in some countries, etc.) are made in auditing practice to promote consensus [Wallage, 1991; Cushing & Loebbecke, 1986]. Our own research examines the degree of consensus about the contents and sequence of the decision-making process, rather than just the consensus as to final results. This is because we wish to find relationships between personal characteristics and auditor's judgment processes.

#### *Auditors' psychological profiles*

There is no doubt that psychological profiles are an important determinant of how humans make decisions [e.g. Pincus, 1990; Bonner & Lewis, 1990, Bonner & Pennington, 1991]. From a theoretical viewpoint, this study may contribute to a better understanding of what factors determine differences in audit judgments. This understanding might eventually lead to judgmental theories of audit planning strategies. Based on such theories, optimal audit plans might be developed. From a practical viewpoint, theories about the judgments which underlie audit planning strategies would be helpful to the technical departments of public accounting firms in providing assessments of the degree to which audit manuals must be structured and adapted to the personal characteristics or psychological profiles of the auditors using the manual.

#### **Research question 3:**

**What factors determine audit planning stage materiality thresholds?**

In order to gather information on the determinants of planning stage materiality thresholds, a case-study was first performed and then a research prototype of an expert system for planning stage materiality judgments was developed. The prototype expert system was tested using both real world and hypothetical cases. Steinbart [1987] asserts that building expert systems is a means for learning more about how humans make decisions. This is the main purpose in this dissertation, although an expert system might also have normative implications, since auditor decision processes might be improved by using such a system [Murphy, 1990]. However it should be stressed that the goal of the current research is not to develop a production expert system, since that would be a complete research project on its own. Rather, the process of developing the system is used to examine auditor materiality and risk assessment knowledge at several levels of abstraction.

#### **1.4 A general audit judgment model**

The relationship between the research questions of this study, and hence the rationale for this study, lies in the general goal: improving audit judgment quality, thus increasing competitiveness in the market for audit services (market share) and

maintaining that market in the long run (total market size). The influence of psychological, demographic and environmental factors on auditors' decision processes can be expressed in the following schematic model (see figure 2).

The research questions asked in this study are incorporated in the model as follows: *Research question 1* (Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?) is represented in the upper left rectangle. This question relates especially to the elements of experience, psychological profile, education and the type of audit firm. *Research question 2* (Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?) is represented by the relationship between the upper left rectangle and the decision process elements in the lower part of the figure. For the purpose of this research question, auditors' problem-solving behavior is defined as: all the cognitive activities which are performed by an auditor in order to solve a certain problem. This would include data search, data- and procedure-storage, information retrieval and decision-making (the latter restricted to decisions about materiality). *Research question 3* (What factors determine planning stage materiality thresholds?) is implicit in the general audit judgment model in figure 2, since the model assumes that the subjects do have expertise in audit materiality threshold assessment. The model describes how this expertise is developed.

### *Expertise and audit decision-making*

In the model, audit expertise is considered to start developing when the auditor makes conceptually similar decisions. For example, an auditor who decides for the first time on the magnitude of a planning stage materiality threshold cannot be considered an expert. However, the same auditor, when he has 10 years of audit experience and hence is experienced in making materiality judgments, could be considered an expert, provided that his cognitive abilities (or, more generally his psychological profile), and the environment in which he gains experience are such that learning from experience is efficient and effective.

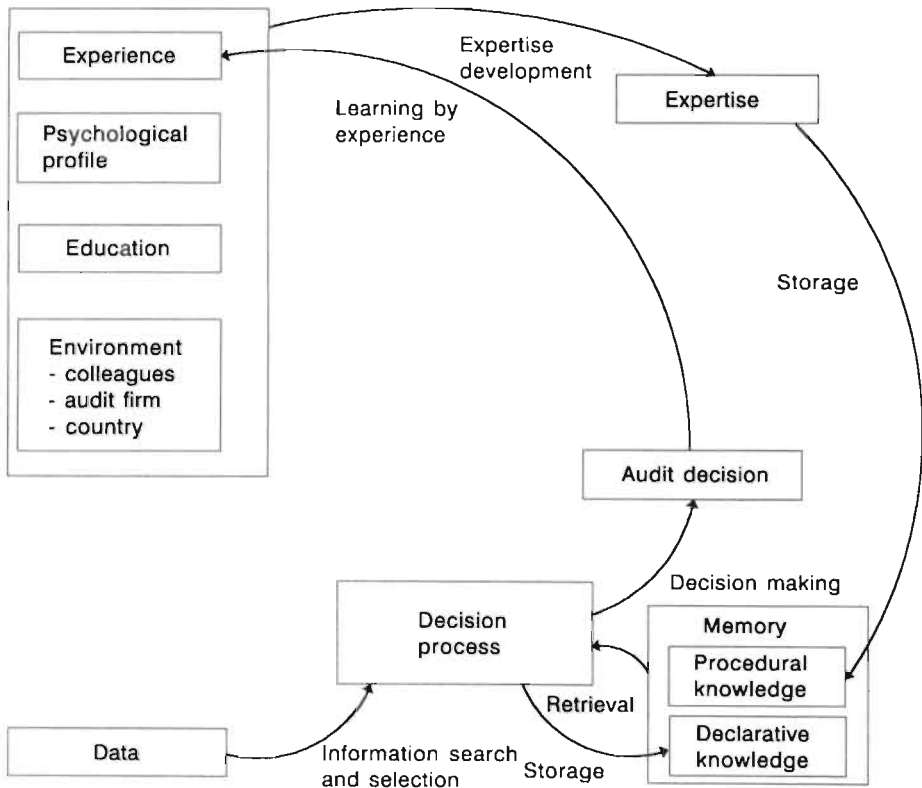
Through repeated judgment and decision making a person learns from experience. By learning from experience, a necessary condition to gain expertise is fulfilled. Experience, personality, and formal education together constitute expertise. In memory this type of knowledge is stored as procedural knowledge. By learning facts from, for example, textbooks, a person stores declarative knowledge in memory. When making decisions, the subject retrieves procedural and declarative knowledge from his memory and combines this knowledge with newly gathered data. Here, the expertise development process is initiated again. So, expertise development is iterative in nature. This conception is clearly reflected in the general audit judgment model in figure 2.

Bedard & Chi [1991] list a number of invariants of expertise in various domains. They include:

- experts know more about their domain than do novices,
- experts' knowledge is better organized,
- experts are better able to distinguish between relevant and irrelevant information,
- in situations where a correct solution procedure exists and is sufficiently understood, experts produce more accurate solutions than novices, although in other situations, such as the materiality and risk judgments in auditing, experts' performance may not be better than that of novices.

These invariants can be recognized in the model: the quantity and organization of knowledge are part of memory, and the issue of relevant and irrelevant information is implicit in the information search activity. The issue of decision quality is of course reflected in the audit decision part of the model.

Figure 2: General audit judgment model.



Bonner & Pennington [1991] investigated why, in some cases, experts perform better and in other cases they perform worse than novices. Two explanations were provided: firstly, there are differences in the knowledge needed for specific tasks, and secondly, there are differences in the types of cognitive processing needed for specific tasks. Bonner & Lewis [1990] and Libby & Frederick [1990] focus on the cognitive determinants of expertise, defined as the innate ability to perform well in certain tasks.

Ashton [1991], and Abdolmohammadi & Wright [1987] view audit experience, which might be an important aspect of auditing expertise, as a concept related to task-specific or domain-specific knowledge. Wright [1988] and Bedard [1991] share this view. Hogarth [1991] also notes the importance of examining the knowledge and processing requirements of different auditing tasks. The same view is held in this study: auditing expertise and related concepts are studied in the specific problem area 'planning stage materiality and risk judgments'.

1.5 Purpose and outline of the study

1.5.1 Issues in audit judgment research

*Research issues*

Three main research questions are asked regarding (1) the relationship between the psychological profiles and demographic characteristics of a population of auditors, (2) the relationship between psychological profiles and demographic characteristics on the one hand and problem-solving behavior on the other hand, and (3) audit expertise in planning stage materiality threshold assessments. In table 1, the research questions are summarized.

Table 1: Summary of research questions of this study.

How are audit planning stage materiality decisions made by auditors?	Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?
	Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?
	What factors determine audit planning stage materiality thresholds?

*Methodological overview*

Three main research approaches have been followed: (1) personality assessment, (2) automated process tracing and verbal protocol analysis, and (3) prototype expert

system development as a means of learning more about human problem-solving processes.

### 1.5.2 Outline of the study

Chapter 2 of the study provides a general overview of the audit judgment literature, in order to place these research and methodological issues in their proper context. In chapter 3, the theoretical basis of the dissertation is developed. Here, the three approaches to the problem of gathering information on how auditors make planning stage materiality and risk judgments are outlined and explained in detail: personality assessment by means of self-assessment questionnaires, verbal protocol analysis and automated process tracing, and expert system development. In chapter 4, the research methodology is discussed. Chapter 5 discusses auditors' psychological profiles in relation to demographic characteristics (research question 1). Chapter 6 deals with the verbal protocol analysis and links it to the demographic variables, to the results of the personality tests, and to the results of the automated process tracings (research question 2). Chapter 7 covers the description of knowledge which is required to make audit planning stage materiality judgments and which is implemented into a *prototype* expert system (research question 3). It should be noted that, although a planning stage risk assessment is not an output of the prototype expert system, a great deal of the described knowledge essentially relates to risk-determining rather than materiality-determining factors. Finally, in chapter 8, the limitations of the study and the theoretical and practical implications of the empirical results are discussed, together with some suggestions for further research.

### Notes

1. Mock & Vertinsky [1985] provide an extensive discussion of audit-related risk definitions.
2. For convenience, detection risk is simplified to consist only of sampling risk, whereas it should also include non-sampling risk. However, this puts no severe restrictions on the interpretation of the sampling model as discussed in this chapter.





---

## An overview of audit judgment research

---

### 2.1 Introduction

The purpose of this chapter is to give an overview of audit judgment research regarding the research questions posed in this study, as it has been conducted in recent decades. This picture will be representative rather than exhaustive. No attempts are made to synthesize the various research findings into one general theory of auditor problem-solving, but a classification scheme is developed which can serve as a basis for presenting audit judgment research findings.

Section 2.2 discusses the various kinds of human judgment which are involved in the auditing process, and explains the need for professional judgment in auditing. Section 2.3 develops a classification system for audit judgment literature. This roughly classifies audit judgment research into two main categories: dynamic and static research. It then becomes clear that only a limited part of the large amount of audit judgment literature is of immediate relevance to this thesis. Dynamic studies regarding materiality and risk are discussed extensively, in order to provide a sound theoretical basis for the current research. The remaining studies, regarding either fields other than materiality and risk assessment, or the static part of audit judgment research, are discussed briefly in order to position the current research.

This chapter is in some senses a complement to chapter 3, which discusses some technical aspects of those research approaches which are particularly relevant for this dissertation. The difference between chapters 2 and 3 can best be expressed by the difference between the questions: '*what* results did we find?' (chapter 2) and '*how* did we find those results, *how* are they related to other research fields, and *how* are they synthesized into general theories of human (c.q. auditor) judgment?' (chapter 3).

### 2.2 Professional judgment in auditing

There is considerable support for the idea that auditing requires substantive human judgment [SAS no. 47]. An audit can be described as a process in which the accounts and the processes underlying them are investigated in such a way that the person who is doing the audit is able to form an opinion as to the fairness of those accounts.<sup>1</sup> Three important elements emerge in this definition: first, auditing is a process, second, auditing is conducted by a person, and third, the result of an audit

is an auditor's opinion. The audit considered as a process implies that a sequence of steps is undertaken toward a certain goal. However, the seemingly structured and programmed audit approaches described in audit manuals contain many options that make their application considerably less objective than one would expect. Each stage has its own specific task elements that require varying skills and knowledge from the auditor. Competence to fulfil a task can only be proved by the auditor's task performance, i.e. during the audit process. The auditor's behavior when solving audit problems may be subject to biases and other personal influences. Libby [1981] asserts that humans, being organisms of limited knowledge and computational capacity, introduce simplifications or heuristics into their choice mechanisms. Research reveals that auditors, like other humans, use these heuristics to a greater or lesser extent depending on their personal characteristics and the nature of the task [see e.g. Joyce & Biddle, 1981a, 1981b]. When forming an opinion about a set of statements, the auditor makes several subjective judgments in each part of the audit process. The auditor integrates these judgments into an overall opinion as to the fairness of the financial statements. The term 'opinion' implicitly indicates a subjective element: professional judgment. This is the most important part of the auditor's decision process because it is the combined result of high level, long-term education and experience, and certain personality structures that together form the basis for expertise in auditing.

Professional judgment is an indispensable element of audit decision making. Auditors rely on professional judgment, implicitly recognizing that professional judgment enables better decision-making than using decision aids in a broad sense (e.g. algorithms, expert systems, detailed audit manuals). However, there is a growing interest within the profession in the development of such decision aids [Gwilliam, 1987]. So there is a paradoxical situation: on the one hand professional judgment is considered to be indispensable, on the other hand auditors seek opportunities to improve decision-making by means of decision aids. The solution to this paradox is that decision aids can partially substitute for professional judgement. If full substitution was possible it would ring in the end of professional judgment in auditing. It would be inefficient to employ professional judgment when some kind of decision aid could do the same job, probably faster and with less likelihood of error. The results of research into the effect of biases caused by heuristics on judgments give an indication of the degree of difference between norm-based judgments (i.e. based on decision aids) and heuristic-based judgments [see e.g. Gibbins, 1977; Uecker & Kinney, 1977; Joyce & Biddle, 1981a, 1981b; Kinney & Uecker, 1982]. The mere existence of audit expert systems indicates that professional judgment in auditing can, at least partially, be captured in computerized models [see e.g. Brown, 1987; Grudnitski, 1986; Peters, 1987; Steinbart, 1987; Dungan, 1983; Biggs & Selfridge, 1986; Dillard & Mutchler, 1986; Hanssen & Messier, 1986; Meservy, 1985]. According to this research, at least partial substitution by such decision aids could lead to a more efficient audit.<sup>2</sup>

The range of possible judgments facing the auditor is wide. Gwilliam [1987] tries to summarize these judgments into four categories: first, prior probability assessments and their revision in the light of new evidence, second, internal control evaluation, third, audit planning decisions, and fourth, materiality decisions. Using this framework of judgment types, the current research focuses on audit risk assessment (as a subset of prior probability assessment) and planning stage materiality decisions, since these two concepts form the base of each audit process model.

### **2.3 A framework for audit judgment literature**

A structured overview of audit judgment literature requires a presentation framework. Several frameworks have been used in (behavioral) decision theory and human information processing research relating to audit judgments in recent decades. Driver & Mock [1975], discuss cognitive domain research, as a special field within the accounting literature on human information processing. They make a distinction between descriptive and normative research. Slovic, Fischhoff & Lichtenstein [1977] classify behavioral decision theory into two main categories: descriptive research and decision aid research. Libby & Lewis [1977] distinguish three categories:<sup>3</sup> lens model studies, probabilistic judgment models and cognitive style studies. Libby [1981] also distinguishes three categories: lens model studies, probabilistic judgment models and, surprisingly, studies of pre-decisional behavior. As could be expected, when Libby and Lewis again collaborated, in 1982, they introduced four categories of human information processing research: lens model studies, probabilistic judgment models, studies of pre-decisional behavior and cognitive style studies. Ashton [1982] makes a distinction between studies within the lens paradigm and studies within the subjective expected utility paradigm. Bedard [1989] distinguishes two main approaches to the study of expertise in auditing: the behavioral and the cognitive approach. The behavioral approach focuses on the output of the decision process and does not dwell on the differences between the cognitive processes of experts and novices, whereas the cognitive approach focuses on these cognitive processes and the knowledge base underlying the behavior of experts and novices. Finally, Johnson, Jamal & Berryman [1989] implicitly classify audit judgment research into four categories: lens model studies, studies of heuristics and biases, studies of decision processes and expert system development.

Some other excellent literature reviews discuss specific problem areas within the field of audit judgment research without explicitly constructing a presentation framework. Colbert [1989] reviews the literature on experience and auditors' judgments in relation to several measures of audit judgment quality, such as consensus, cue weights, self-insight and reliability. Choo [1989] reviews a broad category of judgment literature which has expertise as its central focal point. Davis

& Solomon [1989] provide a review of research on experience and expertise, and give a number of prescriptions for further research on this topic. In their discussion, they support a performance-based notion of expertise, and pinpoint some typical expertise measurement problems.

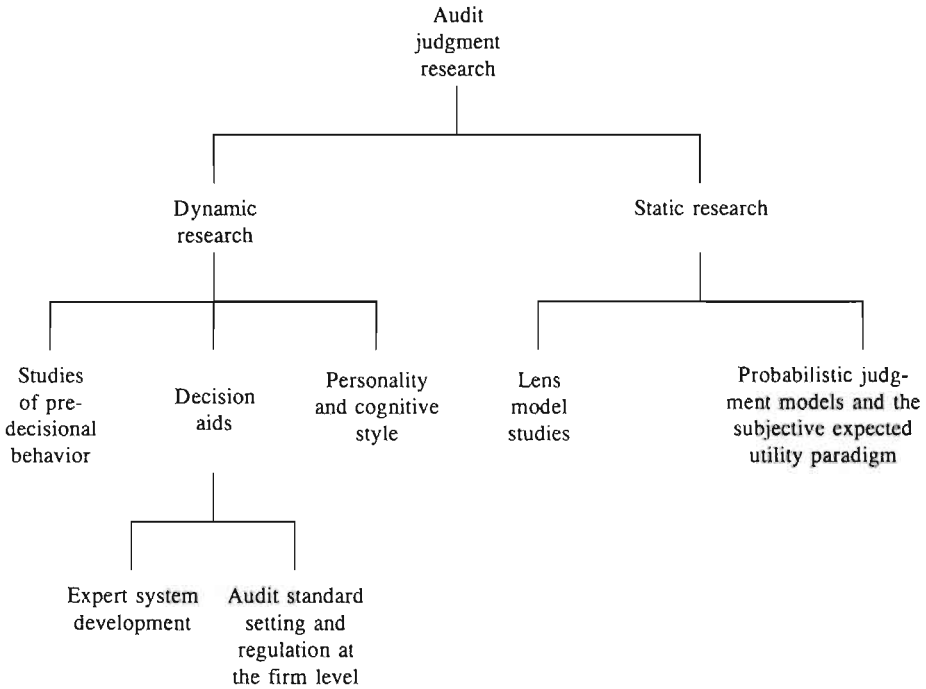
When discussing audit judgment literature, it would seem highly efficient to use one of the most recent classifications as a framework, such as Bedard's [1989] or Johnson, Jamal and Berryman's [1989] schemes. The latter seems the most suitable because it is more detailed and yet is more general, since it is tailored to the whole field of audit judgment and not to some subdivision, such as expertise in auditing. However, the Johnson, Jamal & Berryman classification shares some shortcomings with the other frameworks mentioned. First, it is not clear what criteria have been used to distinguish between the groups. Second, the classifications are presented as consisting of discrete categories, where in fact there is a grey area between the groups. Audit judgment research should be classified along a continuous scale rather than a discrete scale. Third, there are audit judgment studies which cannot be classified within their categories. As an alternative, a classification which is more convenient and applicable to recent research in human information processing and audit judgment can be constructed. In the study of human decision processes, of which the auditor's decision process is an example, there are two main categories: the first involves studying the initial data input of a decision process and the results after processing of the inputs. This is a static approach, or, in Bedard's terms [1989], the behavioral approach. The second focuses on the events between the initial input and the results. This is the dynamic approach, which Bedard calls the cognitive approach. Figure 1 shows the framework that will be used throughout this chapter.

## 2.4 Studies of pre-decisional behavior

Payne, Braunstein & Carroll [1978] describe three methods of studying pre-decisional behavior: verbal protocol analysis, eye movement registration, and explicit information search (i.e. process tracing). In verbal protocol analysis, subjects are normally asked to think aloud while they are solving a case (concurrent protocol) or to recall their thoughts during problem-solving after its completion (retrospective protocol). The verbalized thoughts are registered and analyzed afterwards in order to gather information as to the decision processes of the subjects. Verbal protocols can be used to explore certain problem areas and to formulate hypotheses, to explain behavior during problem-solving, to test hypotheses, and to build computer models of problem-solving behavior (i.e. expert systems). In studies of eye movements in problem-solving, sometimes called REM (rapid eye movement) studies, the information which is, consciously or unconsciously, considered relevant to the problem on hand is traced by recording eye movements. REM studies have not been used in auditing, or even accounting.<sup>4</sup>

Studies of explicit information search, sometimes called process tracing studies, try to break down information processing strategies into (1) cue usage and (2) information access sequence. In order to do so, subjects are required to acquire each information element separately.

**Figure 1:** The audit judgment literature framework.



Studies of pre-decisional behavior have mainly focused on differences between experienced and inexperienced subjects. Some findings were:

- experienced auditors could remember more task elements than did inexperienced auditors, indicating that short term memory functioning is improved by task-related elements which are in long term memory [Weber, 1980],
- experienced auditors employed a systemic information access strategy, whereas inexperienced auditors employed a directed information access strategy<sup>5</sup> [Biggs & Mock, 1980, 1983; Mock & Turner, 1981],
- intermediate and experienced auditors identified the same problem, but intermediates were inclined to increase audit effort more than experienced auditors [Biggs, Mock & Watkins, 1988].

Furthermore, some studies of pre-decisional behavior provide interesting research methodologies, which might be useful in the current study. Williams [1989] employed a computerized information board, on which the subjects could point to items about which they wished to know more.

Some research methodologies which we have categorized under pre-decisional behavior studies (especially verbal protocol analysis) have not been very popular among audit researchers. This is probably due to a seemingly inevitable shortcoming of this kind of research which especially affects its external validity: small sample sizes as a result of its time-consuming nature.<sup>6</sup> However, it must be noted that pre-decisional behavior analyses have been applied in many other fields.

Because the focus in this dissertation is on dynamic research approaches, some technical aspects relating to studies of pre-decisional behavior, and some important historical developments, are discussed in detail in chapter 3.

## **2.5 Expert system development**

Expert system development in auditing has received much attention over the last decade from researchers and practitioners.<sup>7</sup> Excellent overviews of the expert systems that have been developed in accounting and auditing are given by Connell [1987], White [1988], Chandler [1988] and Hayes & van de Poel [1990]. No attempt need be made here to add another overview to this list. However, some recent expert systems will be discussed in depth in appendix C, as they do not appear in any of these studies.

The aim of the current study is not to develop an expert system. However, the expert system development process is used to gather information as to the decision processes of auditors who may be experts in their field. The discussion of expert systems in this section is intended to provide insight into the importance of expert system development for audit judgment research in general. Moreover, the mere existence of expert systems in auditing may be an indication that expertise in auditing, or in some sub field of auditing, does exist. Bedard [1989] provides a critical review of the existing literature on expertise in auditing and concludes that expertise in auditing might exist, dependent on how 'expertise' is defined.<sup>8</sup>

Graham, Damens & Van Ness [1990] developed a rule-based expert system for risk assessment in auditing. Over 100 experts were interviewed in the development process. Knowledge elicitation took place by means of structured interviewing in several stages. In phase 1, an initial prototype was developed and then reviewed by members of the research team in order to enhance it for further testing. In phase 2, a number of audit engagements were examined by means of reviewing key engagement documentation and interviewing engagement team members. Based on

this information, an enhanced prototype was constructed and presented to the key members of the engagement team who gave their comments, in several rounds, on the system, with the aim of adapting it to specific engagement requirements. In phase 3, a field system was constructed for field test evaluation. The field test was conducted in 4 offices and treated twelve different engagements. The engagements were selected to represent a diversity of company sizes and complexities and to test all of the features of the expert system. After each field test session the comments of engagement team members were used to capture additional issues raised during the test sessions. Steinbart [1987] developed a rule-based audit planning stage materiality expert system. This expert system calculates a materiality threshold on the basis of financial and non-financial data. It was found that materiality is determined mainly by the choice of a materiality base and a percentage rate by which that base is to be multiplied. The materiality base is influenced by the nature of the client, future plans of the client, and the auditor's perceptions of the needs of financial statement users. The percentage rate is influenced by the intended use of the financial statements and the nature of the audit engagement.

Since the focus in this dissertation is on dynamic research approaches, some knowledge elicitation techniques and existing concepts of expertise are discussed in detail in chapter 3. However that discussion is aimed at creating a model of expertise in order to study auditors' judgments, rather than at the creation of an expert system for its own sake.

## **2.6 Audit standard-setting and regulation at the firm level**

Audit standard-setting is the process of issuing statements and guidelines about the standards which auditors must or may comply with when conducting an audit. On an international level these standards are issued by regulatory bodies such as the IAPC of the IFAC. On a national level standards are issued by professional organizations such as the AICPA, the ICAEW and NIVRA.<sup>9</sup> On a micro level, these standards are issued by the technical departments of auditing firms. The audit manuals developed by these firms contain detailed prescriptions and decision aids about how to conduct an audit in the most efficient manner. The international and national standards and developments are incorporated into these micro level prescriptions.

In our general model of audit judgments (chapter 1, figure 2), audit standard-setting and regulation at the firm level are represented by the environment in which audit judgments are made. Professional and firm level standards, together with client and user characteristics, constitute the context of any audit judgment. So audit standard research is directly relevant only to the contextual part of our model. However, the technical departments of audit firms might improve their standard-setting process by gaining an understanding of what personal variables (in the

model: expertise, experience, personality, and educational background) determine the degree to which there is consensus among auditors on decision processes and the outcomes of decision processes. This aspect of the current research makes it interesting to review the literature on audit standard-setting and regulation at the firm level.

Audit standards research is practice-oriented by nature. Judgment research in this field focuses on the study of decision-making regarding the quantitative or qualitative values of the various components of the audit risk model, as this is prescribed in a particular audit manual. Brown [1962] emphasized the importance of quantifying an internal control questionnaire. Such a questionnaire is used by auditors as a checklist of potential internal control measures. It consists of a number of questions that have been assigned weights. Each question must be answered with 'yes' or 'no'. The quality of the internal control system is measured by the sum of the weights of those questions to which the auditor answered 'yes', divided by the maximum possible score (effectiveness index).<sup>10</sup> Although the method is appealing, because of its simplicity, some criticisms must be noted. First, the process of assigning weights to questions in the internal control questionnaire is a matter of professional judgment. So what really happens is that the subjective element in the internal control evaluation is moved from the fieldwork stage to the preparation stage (i.e. the preparation of the internal control questionnaire). Because of the need for consistency in the effectiveness index over time, weights cannot be adjusted, resulting in a static, unrealistic reaction to internal control developments. Finally, the algorithm for the determination of the effectiveness index does not contain a correction for synergetic effects. Positive answers to four 1-point questions cannot generally be considered to have the same influence as a positive answer to one 4-point question. However, despite this criticism the approach set the scene for a more quantitative audit approach which is now commonplace, but which was almost revolutionary in the sixties.

Murphy [1990] examined the effect of using an expert system on the development of expertise among novices. The subjects were 67 senior level accounting students. Three treatment groups were distinguished: one was presented with an expert system with explanations, one with an expert system without explanations, and one with a non-automated practice aid. Three hypotheses were developed: first (H1), there will be no difference across treatments in subjects' SFAS91<sup>11</sup> post-test scores, second (H2), there will be no difference across treatments in subjects' ability to classify the proposed accounting treatment of loan fees and costs as correct or incorrect, and third (H3), there will be no difference across treatments in subjects' ability to classify the proposed accounting treatment of lease costs as correct or incorrect. H1 and H2 could be rejected: subjects who did not use an expert system had higher scores than expert system users who received explanations in the semantic memory development task (understanding the provisions of SFAS91) as



well as in the combined semantic/episodic memory development task (a loan decision case). Subjects who did not use an expert system had higher scores, in the semantic memory development task alone, than expert system users who did not receive explanations. H3 could not be rejected: no differences could be observed in the episodic memory development task (a lease decision case).

## **2.7 Personality and cognitive style**

In cognitive style research, variables regarding personality and cognitive structure are investigated in order to find optimal relationships between information reports and information processing, i.e. individualized information systems.<sup>12</sup> The processes of audit standard-setting and regulation at the firm level might be improved by gaining an understanding of auditors' personality structures and by tailoring audit manuals to the information processing capabilities of auditors. If there are differential performance peaks, i.e., some subjects perform better when provided with certain types of information, audit manuals could be tailored to the specific information processing capabilities of auditors to assist each auditor to achieve his optimum performance.

Benbasat & Dexter [1979] based their research on the values and events hypothesis as developed by Sorter [1969].<sup>13</sup> They found differential performance peaking, but due to their limited sample size the results were less generalizable than those obtained by Lusk [1979], who found that differential performance peaking did not exist. On the other hand, Benbasat and Dexter's approach - investigating the values and events hypothesis - is appealing. Pincus [1990, 1991] categorized auditors on the basis of an embedded figures test, a tolerance for ambiguity test, and a category width test (risk-taking propensity),<sup>14</sup> as well as experience in auditing. Of the various results that were obtained, the findings relevant to the current research are that those with high tolerance for ambiguity more often judged accounts not to be materially misstated, and that highly experienced auditors were more confident in decisions leading to an unqualified opinion and less confident in decisions leading to a qualified or an adverse opinion.

Because, the focus in this dissertation is on dynamic research approaches, personality and cognitive style research and the techniques used in gathering information regarding personality and cognitive style are discussed in detail in chapter 3. In that chapter, a number of cognitive style and personality test instruments are discussed, as well as some typical characteristics of the scores which have been obtained.

## **2.8 Lens model studies**

The lens model approach, as presented by Brunswik [1952], offers a static

framework for psychological research, in that a set of cues relevant to a certain decision is identified. In the model regression equations, discriminant functions, and correlation measures as well as ANOVA techniques are used to describe the relationships between events and cues, between cues and decisions, and between decisions and events.<sup>15</sup> Since, in judgmental decision situations there is uncertainty about the decision outcomes [Sniezak & Henry, 1989], the decision-maker has no clear picture of the event to be judged. Hence, the relation between cues and events is probabilistic.

For example: an auditor must determine a planning stage materiality threshold based on information about client-specific circumstances. Here the decision is the planning stage materiality threshold, the cues are the client-specific circumstances, and the criterion event is the actual materiality level at which the reader of the financial statements will change his decision. In this specific situation the auditor is separated from the criterion event by time and space. He is unable to determine what decisions, and by what decision-makers, will be made since decisions are taken in the future (time) and the decision-makers are anonymous (space).

Messier [1983] mentions four frequently-used standard criteria to measure quality of professional judgments: consensus, cue weighting, self-insight, and reliability (i.e. stability). Most lens model studies focus on the experience effects in relation to these decision quality criteria [e.g. Ashton, 1974; Moriarity & Barron, 1976; Joyce, 1976; Moriarity & Barron, 1979; Ashton & Kramer, 1980; Ashton & Brown, 1980; Mock & Turner, 1979, 1981; Gaumnitz et al., 1982; Hamilton & Wright, 1982; Krogstad *et al.*, 1984; Frederick & Libby, 1986; Abdolmohammadi & Wright, 1987; Libby & Frederick, 1990; Moeckel, 1990, 1991; Bedard, 1991].

Moriarity & Barron [1976] investigated consensus in relation to materiality determining factors. It was found that subjects were consistent regarding net income as an explanatory variable of the height of materiality thresholds. However, there was no consensus regarding the effects of earnings trend and company size on materiality judgments. Moriarity & Barron replicated their 1976 study in 1979. Again, in a materiality assessment task subjects exhibited hardly any consensus.

**2.9 Probabilistic judgment models and the subjective expected utility paradigm**  
Another method which has frequently been applied in accounting research on the utilization of information by humans focuses on the assessment and revision of probabilistic judgments about uncertain future events. Revisions of probabilities are studied in relation to the normative implications of Bayes' theorem.<sup>16</sup> Peterson & Beach [1967] concluded that observed human behavior closely approximated the normative Bayesian model. However, contrary to this finding, later research results reveal that revisions are generally incomplete due to excessive conservatism [Slovic

& Lichtenstein, 1971]. This raised the question of how this phenomenon could be explained. Tversky & Kahneman [1974] introduced three heuristics employed by decision-makers due to their cognitive limitations: representativity, availability, and anchoring and adjustment. Subsequent studies resulted in considerable extensions of this list [Hogarth, 1980]. Van de Poel [1986] remarks that the long lists of heuristics that have been observed empirically mirror the difficulties faced by accounting researchers in trying to arrive at some basic theory which serves as an alternative for subjective expected utility theory. Expected utility theory has been widely accepted as the normative model of rational choice [Van de Poel, 1986]. Rational choice models are based on the assumption that decision-makers are able to specify all possible outcomes of all possible actions and know their (stable) preferences. Uncertainty is incorporated into these models by means of subjective probability distributions. However, despite these unrealistic assumptions, normative decision theory is widely utilized in auditing and other accounting sub-fields. The auditor is thereby assumed to be a Bayesian processor of information.

Lens model studies and probabilistic judgment studies exhibit a certain overlap. In the classification of audit judgment research made in this section a decision-theoretic, or Bayesian approach to the study of auditors' problem-solving is followed. Studies that involve the determination of relationships between personality or experience on the one hand and consensus, self-insight or consistency on the other hand are labelled lens model studies or cognitive style studies.

Newton [1977] investigated the decision processes of auditors when resolving materiality issues in relation to beta risk assessments. It was found that the more averse the auditor is to accepting beta risk, the less uncertainty he will accept, and the lower his materiality threshold will be. This finding is especially relevant for the current study since materiality and risk are also being studied simultaneously.

Libby [1985] tested for the use of the availability heuristic regarding the occurrence of errors. It was found that availability and perceived frequency of errors were positively correlated. However, there were no recency effects, indicating that availability did not completely influence auditors' decision-making. Mock *et al.* [1993] found that Bayes' theorem was only partially followed by auditors, a conclusion that is in concordance with earlier research findings.

An important conclusion from the results of the probabilistic judgment studies as described in this section, is that the normative implications of Bayes' theorem are only partially valid for auditors. Some heuristics may be used (availability, anchoring and adjustment, representativeness) in order to simplify cognitively complex matters. However, research provides mixed results regarding the use of

heuristics by auditors. These mixed results could be due to differences in task environments.

### **2.10 Positioning the current research relative to previous research**

As can be observed in appendix C, audit judgment studies of pre-decisional behavior focus primarily on assessing the differences between novices and experts or between inexperienced and experienced subjects. In some tasks, experience appears to be the most important determinant of expertise. For example: auditors remembered more EDP controls than students, experienced auditors performed a systemic information access strategy whereas inexperienced auditors performed a directed information access strategy, experienced auditors were more able to design proper audit procedures than less experienced auditors, and experienced auditors more often discovered management fraud. In other audit tasks, hardly any experience effects were found.

Studies of audit judgment decision aids focus primarily on improvements or deteriorations in performance as a result of using the decision aid. The effects found are highly ambiguous. For example, from one study [Weber, 1978] it appears that auditors using a simulation decision aid are more accurate, have more confidence in their own decisions, are faster, and are more satisfied with the audit evidence. In another study, Murphy [1990] found that using a decision aid leads to inferior learning processes. Furthermore, Mock & Turner [1981] found that most auditors do not increase sample sizes when decision aids are used, whereas Kachelmeier & Messier [1990] found the opposite result.

Studies of audit judgment cognitive styles focus primarily on the relationships between scores on self-description tests and actual performance in auditing tasks. Classifications coming from the Myers-Briggs Type Indicator, MacDonald's tolerance for ambiguity test, the embedded figures test, risk-taking propensity tests, and the decisive, hierarchic, integratives, complex, and flexible typology as developed by Driver & Mock [1975], are used to investigate whether certain subjects perform better when provided with certain types of information (differential performance peaking). Results regarding differential performance peaking in an audit environment are somewhat ambiguous. However, relationships between self-insight test scores and actual performance can be observed over a wide variety of tasks and tests.

Lens model studies of audit judgments focus on the consensus among auditors with varying levels of experience, in order to assess decision quality more objectively. Generally, in internal control tasks it appears that auditors rate internal controls uniformly [Gaumnitz *et al.*, 1982]. However, if internal controls are strong, experts exhibit more consensus than when internal controls are weak [Bedard,

1991]. Experts have a better understanding of relations between internal control quality and error probabilities [Frederick & Libby, 1986]. Experienced auditors have a greater tolerance for higher risk situations than less experienced auditors [Abdolmohammadi & Wright, 1987] and more often integrate different pieces of information correctly [Moeckel, 1990]. In general it can be concluded that experience and decision quality are positively correlated. An exception is found in Hamilton & Wright [1982], where experience and consensus are negatively correlated. Audit judgment studies of probabilistic judgment models focus on consensus and auditors' use of heuristics. Auditing experience and statistical background (education) does not influence prior probability assessments, auditors are not Bayesian when determining sample sizes, in some tasks auditors employ the anchoring and adjustment heuristic unjustly (fraud), in other tasks the anchor has no significant influence on decisions (internal control rating, audit reporting). In materiality, fraud detection, internal control and reporting tasks, consensus is low.

Three overall conclusions can be drawn from previous research which are relevant to the research question of this study:

- audit expertise is mainly determined by the number of years of experience in specific audit tasks, but the way in which audit expertise is defined determines to what degree audit expertise exists,
- consensus is prevalent in certain tasks and absent in other tasks, i.e., it is task dependent,
- personality assessments can provide good predictions of actual problem-solving behavior.

Bearing in mind that this overview of the literature might be incomplete, we have found six studies which are directly relevant for the research questions of this dissertation. These are (in chronological order):

Biggs, Mock & Watkins, 1988;  
Williams, 1990;  
Mock, Wright, Washington & Krishnamoorthy, 1990;  
Pincus, 1990;  
Graham, Damens & Van Ness, 1990; and  
Steinbart, 1987.

Each of these studies treats materiality and/or risk within a dynamic research methodology.

In chapter 3, some technical aspects of the research approaches which are particularly relevant for this dissertation are discussed in detail. Chapter 4 will discuss our research methodology, and chapters 5, 6 and 7 will investigate the findings from audit judgment literature in general, to assess whether or not they are confirmed by the current research. In these analyses, use is made of the techniques and methodologies as discussed in chapter 3.

## Notes

1. This definition differs from that published by a committee of the American Accounting Association in 1972: 'Auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users'. The AAA definition is less specific and provides for a broader view of auditing services [Kinney, 1987]. However, the definition used in this thesis fits well with the definition of the AAA [1972], and since it is more specific, it will be used instead.
2. Although, to the author's knowledge, no study aimed at finding the optimal relationship between decision aids and professional judgment has been done, and there has been only limited research which even gestures in this direction [Weber, 1978; Ebert & Kruse, 1978], we can say that, if there is a non-proportional trade-off between professional judgment and decision aids in terms of the decision quality resulting, then there must be an optimum combination of professional judgment and decision aids.
3. In Feenstra [1988] and Limperg Instituut [1990] it is suggested that the term category should be substituted by the term paradigm, in order to indicate that categories in this meaning are merely modes of thought.
4. Russo [1978] made an eye movement study in marketing.
5. The terms derive from Biggs & Mock [1980, 1983]. A directed strategy involves a separate search for information needed for each of (four) decisions to be made. A systemic strategy involves building up an overall picture of the problem field and subsequent decision-making for all (four) problems simultaneously.
6. I would suggest that the analysis of pre-decisional behavior as a research methodology must be conducted following the same statistical standards as are used in other research methodologies. Therefore, results from small sample studies must be interpreted carefully because of a potential lack of external validity.
7. Auditing firms have put great effort into the development of expert systems for auditing purposes. Unfortunately, hardly any publications have come to light about these systems. However, based on a snapshot study of audit approaches, a few examples of expert systems can be given. First, the construction of audit programmes tailored to client-specific circumstances. This system is in use but produces large amounts of paper, which causes some resistance among potential users. Second, audit programme development. This system is only in use for a limited number of industries. Third, inherent risk assessment. The system consists of four modules at varying stages of development: management (field tested), audit objectives (adjusted prototype), product-market aspects (prototype), and financial aspects (prototype in development).
8. However, since most audit decisions are judgments, and hence cannot be objectively classified as good or bad, there could be expert systems which are in fact not expert systems but rather knowledge systems. These systems would contain knowledge bases, containing the knowledge of experienced auditors who are in fact not experts.

9. The AICPA (American Institute of Certified Public Accountants), the ICAEW (Institute of Chartered Accountants in England and Wales), NIVRA (Nederlands Instituut van registeraccountants) and other national organizations of auditors have as their main goal the promotion of the interests of their members. Standards are issued as part of this goal.

10. If this questionnaire were computerized it would be considered an expert system.

11. SFAS91 (Accounting for Nonrefundable Fees and Costs Associated with Originating or Acquiring Loans and Initial Direct Costs of Leases) was chosen because it provided a structured, deterministic, and reasonably complex experimental task.

12. Libby [1982] asserts that this kind of research can only have relevance in personnel selection, because individualized information systems require different hardware and software systems for different groups of users. This is difficult to realize because of problems in identifying the interactions between system type and decision quality and because of the high costs of such systems.

13. Since the value and events hypothesis [Sorter, 1969] has been developed for accountants in general, it is also applicable to auditors. The value approach assumes that information needs are known and specified so that accounting theory can be used to deduce input values that can produce optimal results within the context of certain known decision models. The events approach assumes that the decision models of users of information need not be known and that the accountant can only play a role in providing information which might be useful in a variety of potential decision models. In the latter approach, highly aggregated reports are not appropriate.

14. These tests, among others, will be discussed in chapter 3.

15. Regression equations and discriminant analysis are used when cues can be measured on a continuous scale. ANOVA is used when cues are considered to be factors which can be divided into a few discrete categories.

16. This theorem is named after its discoverer, Thomas Bayes, an English clergyman who lived from 1702 to 1761. It states that prior probabilities will be revised to posterior probabilities when new information induces this in conformity to a rigid mathematical model. Suppose that an individual forms a subjective view of the probability that event B will occur: this prior probability is  $P(B)$ . If subsequently a new piece of information becomes available, e.g. that event A has occurred, then this may cause a modification of the judgment about the probability of event B occurring. Since A is known, the probability of B is a posterior probability:  $P(B|A)$ . The relation between these probabilities can be stated as:  $P(B|A) = \{P(B|A)P(B)\}/P(A)$ .





## **Technical aspects of personality assessment, verbal protocol analysis, and expert system development in relation to audit judgment**

---

### **3.1 Introduction**

In order to find answers to the research questions posed in this dissertation, we have chosen a dynamic research methodology, which is aimed at directly studying auditors' decision processes rather than just the input and output. Chapter 2 has provided an overview of audit judgment research in recent decades. In that chapter, some *findings* of audit judgment research are discussed, distinguishing between dynamic and static research. Having chosen a dynamic research methodology for the gathering and analysis of empirical data, three dynamic *approaches* to the research questions are outlined and explained in this chapter: personality and cognition (§3.2), verbal protocol analysis including automated process tracing (§3.3), and expert system development (§3.4). These research approaches are applied to the data gathering and analysis in this study. The discussion in this chapter is therefore of a technical nature, since an application of specific research methodologies requires detailed knowledge of the techniques that are commonly used in those methodologies.

### **3.2 Personality and cognition**

Cognitive style may be considered a subset of personality structure. Cognition refers to the concept of learning, more specifically to the perceptual and intellectual aspects of learning. Cognitive style thus refers to how humans learn from formal education and experience. Cognitive style determines the way in which an individual processes, transforms, and restructures stimuli from the environment (information), thus shaping the individual's response. Doctor & Hamilton [1973] provided a definition of cognitive style which has been widely used: cognitive style is 'the characteristic, self-consistent way of functioning an individual exhibits across perceptual and intellectual activities'. Cognitive style, in this view, is a narrower concept than personality, since personality can be defined as the combined effect of sundry traits such as emotional stability, cautiousness, vigor, ascendancy and responsibility [Macintosh, 1985].

Cognitive style research, as a subset of personality research in accounting, has received a great deal of attention because of its close relation to information acquisition and processing studies [Libby & Lewis, 1982]. Cognitive style theory starts from the premise that the broad dimensions of a person's functioning (e.g.

decision behavior) may be determinable from his cognitive style. Measurement of cognitive style is thus potentially a key to tailor-made information systems and more efficient reporting in general.

A number of cognitive style measurement instruments have been developed in the application of cognitive psychology to accounting [Keen & Bronsema, 1981]. Three kinds of self-concept questionnaire-based cognitive style tests can be distinguished: embedded figures tests, Jungian typology tests (usually referred to as Myers-Briggs type indicators), and converger-diverger tests. A fourth category distinguished by Keen & Bronsema [1981], cognitive complexity theory, uses performance rather than self-concept as the basis for classification. Since this approach in its pure form is more closely related to process tracing studies, it will not be discussed in this section. However, decision style theory, a combination of the converger-diverger construct and cognitive complexity theory, is added to the Keen & Bronsema classification.

Waller & Felix [1984] describe the cognitive process by which auditors learn from experience. Their study is particularly relevant for research into auditors' cognitive styles because it discusses experience effects in relation to several concepts of learning. However they do not make the link between cognitive style assessment and learning from experience, of which cognitive style is an indicator. Such a link, if it could be established, holds out the promise of a well-grounded theory of cognition in relation to experience in auditing. However, this challenge has not been taken up by audit researchers, probably due to the implicit assumption in cognitive style theory that cognitive style, considered to be part of one's personality, is invariable over time and in any context.

Criticism of the methods and techniques used in cognitive style research, and in personality research in general, focuses on whether the variables measured are valid representations of stable personal characteristics, as the Doctor & Hamilton definition maintains, or mere dispositions that refer to tendencies to respond in a certain way under certain circumstances. Van de Poel [1986], in an experiment among 45 Dutch graduate students and 23 experts from the field of accounting, found no significant relationships between decision confidence, Myers-Briggs type indicator, tolerance for ambiguity, and performance evaluation style. To explain these results, he argued that the 'disposition' argument must be valid, since subjects did not exhibit the assessed personality structures under laboratory conditions. A second argument, which is difficult to verify, is that the measurement instruments (i.e. the cognitive style questionnaires) were imperfect and hence are to be rejected.<sup>1</sup>

Glover *et al.* [1990] describe three distinctions regarding memory: short-term versus long-term memory, episodic versus semantic memory and declarative versus

procedural knowledge. Newell & Simon [1972] define short-term memory as the information content that is directly available for cognitive processing at a given moment. Short-term memory can be compared to the working memory of a computer in that its capacity is very limited, comprising only the knowledge currently in use. Research [Simon, 1974] has indicated that the capacity of short-term memory is only five to nine chunks of information, independent of task characteristics. Long-term memory is the information content stored for later use and only retrievable via short-term memory. Here again an analogy with the computer can be made,<sup>2</sup> in that long-term memory has a much greater capacity than short-term memory. However, human long-term memory has no practical capacity constraints. Episodic memory is the reflection of a bottom-up learning process containing autobiographic experiences, whereas semantic memory is the reflection of a top-down learning process containing general knowledge of concepts and meanings.

The question as to the organization of long-term memory is often illustrated by means of the dichotomies episodic-semantic and declarative-procedural. An understanding of learning from experience requires an explication of semantic memory structures and of the interaction between semantic and episodic memory. Moreover, the differentiation between declarative and procedural knowledge is especially relevant for the organization of long-term memory because long-term memory may contain 'knowing that' information (declarative) and 'knowing how' information (procedural). Despite the seemingly loose relation to cognition, both knowledge components refer to the process of learning, although it is to facts learning on the one hand and skills learning on the other hand. Declarative knowledge is organized into long-term memory by means of categories and schemata. The distinction between categories and schemata is not clearly understood [Waller & Felix, 1984]. This would indicate that the difference is gradual: categories, in this sense, organize knowledge on the basis of class membership or similarity,<sup>3</sup> while schemata, in contrast, organize knowledge on the basis of generalized cognitive structures which represent knowledge about how the world works. Schemata adapt to experience via assimilation and accommodation. Procedural knowledge is organized into long-term memory by means of production systems (i.e. if..then-statements). Table 1 represents this classification of (long-term) memory and knowledge components.

### 3.2.1 Auditors' accounting and auditing knowledge

The auditor, when forming an opinion as to the fairness of financial statements, must consider materiality *in auditing*. However, since in the compilation of financial statements various decisions are made as regards presentation and disclosure, there is also another materiality issue: materiality *in accounting*. Leslie [1985] argues that accounting and auditing materiality involve the same object (i.e. the financial statements). So it is obvious that both must have the same magnitude.

Moreover, the auditor gives an opinion as to the fairness of financial statements, meaning that he declares that these statements are correct within certain boundaries. When a client's management prepares financial statements in such a manner that a fair representation of the firm's economic position and performance is given, the auditor should give an unqualified opinion. However, the question as to the definition of fairness can only be answered subjectively. The auditor must therefore choose materiality thresholds such that management's concept of true and fair is unjustly refuted or accepted in only a limited number of engagements. Ideally, the two materiality thresholds should converge to a level at which auditor and management agree.<sup>4</sup>

**Table 1:** A classification of long-term memory and knowledge components.

Long-term Memory (LTM)					
Episodic memory			Semantic memory		
Procedural knowledge	Declarative knowledge		Procedural knowledge	Declarative knowledge	
Production systems	Categories	Schemata	Production systems	Categories	Schemata

The inexperienced auditor's accounting and auditing knowledge, used in arriving at an opinion, is purely declarative because he knows in theory what accounting and auditing rules to apply. Later on, when the auditor becomes more experienced, a transformation of his declarative knowledge into procedural knowledge takes place: he knows how to apply his declarative knowledge. The difference between declarative and procedural knowledge can best be expressed by the difference between knowing that (declarative) and knowing how (procedural). So, gaining professional experience goes together with enlarging the contents and number of production systems in memory. This adaptation of the auditor's procedural knowledge is likely to interact with schematic memory structures which represent concepts such as user relevance, materiality and fairness of disclosure. An interaction with category memory is initiated as the auditor learns to apply knowledge gained with client X to the engagement with client Y, or in other words, when the auditor is able to classify clients into his own (mental) categories.

Formal education leads to an adaptation of semantic memory, whereas experience leads to an adaptation of episodic memory. This notion can be exemplified by means of the auditor's opinion formulation process. At first, the novice auditor learns, through formal education, what sorts of opinions can be issued and in what circumstances each of these is appropriate. This knowledge is typically concept-

driven or based on top-down learning, and is therefore stored in (long-term) semantic memory. Part of this knowledge is declarative (e.g. knowledge about client-specific circumstances and the set of production rules which must hold to issue a certain opinion), and part is procedural (e.g. the skill of identifying client-specific circumstances which require a certain opinion, a question of correctly applying declarative knowledge about production rules). Later on, the novice auditor gains experience by means of bottom-up learning, thus extending his episodic memory. Here again, part of this knowledge is declarative and part is procedural.

### 3.2.2 Personality tests

When trying to find significant relationships between personality (or more specifically cognitive style and decision style) on the one hand, and all kinds of performance indicators on the other [see e.g. Driver & Mock, 1975; Weber, 1978; Lusk, 1979; Benbasat & Dexter, 1979; Pincus, 1990, 1991], psychometric measures need to be developed. In psychological research a number of cognitive style and other personality tests have been applied in recent decades. In this section some widely applied tests are discussed (references are to the landmark papers):

- the Myers-Briggs Type Indicator (MBTI) [Keen & Bronsema, 1981],
- the Driver & Mock decision style test [Driver & Mock, 1975],
- the Embedded Figures test [Witkin, 1950],
- the Analytic/Heuristic test [Mock *et al.*, 1972],
- the Tolerance for Ambiguity test [MacDonald, 1970],
- the Hopwood leadership style test [Hopwood, 1972],
- the Seiler & Bartlett leadership style test [Seiler & Bartlett, 1982],
- the House *et al.* leadership style [House *et al.*, 1990].

One of the most popular cognitive style tests is the Myers-Briggs type indicator (MBTI). Research with respect to the Myers-Briggs type indicator has traditionally been conducted within the field of Management Accounting. Examples are found in Henderson & Nutt [1980], Kerin & Slocum [1981], Schweiger & Jago [1982], Blaylock & Rees [1984], Nutt [1990], Ruble & Cosier [1990], and Davis *et al.* [1990]. Jung [1921] held that the wide variation in human behavior he observed in his psycho-analytical practice could be traced back to four dimensions on which humans differ with regard to their mental functioning: cognitive reality, cognitive mode, cognitive process for perceiving, and cognitive process for judging. *Cognitive reality* refers to the extent to which people prefer the external world of people and events over the internal world of archetypes, ideas and concepts. This distinction leads to the dichotomy: extraverted-introverted. *Cognitive mode* refers to the extent to which people prefer perceiving and understanding over judging and vice versa. This distinction leads to the dichotomy: perceiving-judging. *Cognitive process for perceiving* refers to the extent to which people rely on facts, concrete

data and actual experience as opposed to concepts, theories, ideas and hypothetical relationships when perceiving. This distinction leads to the dichotomy: sensing-intuition. Finally, *cognitive process for judgment* refers to the extent to which people base their judgments on personal values and feelings as opposed to impersonal analysis and logic. This distinction leads to the dichotomy: feeling-thinking. The dimensions of this classification which apply to perceiving and judging are especially relevant for behavioral accounting research since accounting deals with the acquisition (perception)<sup>5</sup> of information and the processing (judgment)<sup>6</sup> of information. For that reason the cognitive style pairs sensing-intuition and feeling-thinking will be discussed in more depth, presenting a set of typical characteristics for each category [Macintosh, 1985].

People who have a strong preference for a sensing way of perceiving become aware of their environment directly through the five senses. They have a strong preference for facts and actualities, and hence they are prone to gather as much evidence as possible when trying to understand things. They do not skip pieces of text when reading, indicating a certain carefulness and accuracy. When calculating they take much time. Learning often takes place by rote. Furthermore, they have problems in making abstractions and in using symbolic representations. Intelligence test are difficult for sensing types. They do not trust indirect sensing via the unconscious or via direct sensing made by other people. They are friendly users of the detailed information provided by accounting information systems. They enjoy problem-solving, but only go as far as the actualities and the facts gathered allow, not extrapolating to the future or to other situations. When using information systems they require some sort of coaching (human or computerized, in the form of decision-support systems) since they cannot cope with the symbolic and abstract representations that are part of many information systems. The latter is somewhat paradoxical: sensing types need lots of information, preferably detailed, but are not capable of using so much information efficiently.

Intuitive types, as opposed to sensing types, are more taken with possibilities than with facts and actualities. They admit only the sense impressions which relate to the inspiration of the moment to consciousness. The known and established is used only to arrive at an advanced point, leaving the intermediate steps to the unconscious. Intuitive types perform independent of their physical surroundings. They are inventive and original, but are not careful observers. They are initiators, inventors, promoters and entrepreneurs. Their need for achievement is very high. Little attention is paid to the present: the future is the only thing that counts. They see objects and events in their totality with properties not derivable from their parts. Since they are not really interested in accounting information, they are not good users of formal information systems.

People who have a strong preference for a thinking way of judgment rely on

impersonal, logical thinking, ruling out as much as possible their subjective feelings. They value logic and objectivity more than sentiment. However, their thinking is not necessarily sound, possibly relying on ambiguous information as if it were true. They act businesslike and are brief in their communication, even lacking a certain friendliness. As a result they have more executive than social ability. Judgments are made on the basis of a true-false criterion, valuing absolute truth more than tact. Furthermore, they like categorizing. Given these characteristics it is not strange to observe the thinking type having a propensity to over-utilize models and rules and being the best and most friendly users of information systems.

Feeling types, as opposed to thinking types, prefer a personal, subjective approach to life. They are tender-hearted and naturally friendly. As a result, they are good at establishing and maintaining human relations. However, they are prone to suppress and undervalue thinking, especially when it is in conflict with their feelings. They find it difficult to know where to start a statement or in what order to present their views. For that reason, they give more detail than strictly necessary, repeat themselves and ramble. Sometimes this behavior is considered very inefficient. Their judgments are based on the agreeable-disagreeable perspective. Because feeling is so important, they are not effective users of information systems that produce formal accounting reports. In addition, they are interested in information in artistic forms such as poetry or paintings.<sup>7</sup> When compared to thinking types they are better at using graphic reports than tabular reports.

The MBTI is assessed by means of a self-description questionnaire in which respondents are asked to answer a number of questions.<sup>8</sup> Every question has only two alternative answers, each relating to one of two cognitive styles. There is no time-limit as to the completion of the questionnaire.

Some personality tests, unlike the MBTI, have been used in accounting and auditing research. Driver & Mock [1975] developed a decision style model based on two dimensions: the amount of information used (minimal versus maximum) and degree of focus (multiple solutions versus one solution). In a way, an analogy can be made with the perceiving and judgment dimensions of the Jungian typology, since sensing and intuitive types mainly differ as to the amount of information used, while feeling and thinking types differ in their preference for agreeable or true solutions (or disagreeable versus false solutions).<sup>9</sup> The decision style model identifies four typical modes of decision-making: a flexible style, an integrative style, a decisive style and a hierarchic style. The *decisive* individual uses a minimal amount of information and likes the information to generate only one solution. He has a high concern for speed, efficiency and consistency. Short range planning, based on little data, is combined with a tight control of results. His goals are those of the organization, in order to avoid extensive discussions on the point. The

decisive personality prefers an hierarchic organization with clear-cut rules and a relatively small span of control. In accounting he prefers short, summarized reports focusing on results and actions. The *flexible* individual, like the decisive, uses minimal data. However, he ascribes different meanings to that information at different times. Since intuition is a very important aspect of this style, planning is not considered to be meaningful. There is hardly any reliance on organizational goals because this would limit the necessary flexibility. Instead the flexible personality pursues self-oriented goals. Also, a loose organization is preferred, hopefully resulting in confusion among subordinates, because this will ultimately lead to creativity and control by the flexible manager. Finally, the flexible type likes variety and many, however briefly stated, solutions so that he himself can make a comparison. The *hierarchic* individual uses masses of data aimed at reaching the one best solution. Long range planning is considered very important. Control over results and methods is sought in trying to meet the few personal goals. The hierarchic favors the classic organization with much attention being paid to formal procedures, and with broad spans of control. Brief or inconsistent reports are not tolerated, instead long, formal and thorough reports are favored in communications. The *integrative* individual, like the hierarchic, uses masses of data. However, he will generate a multitude of solutions at the same time. Integratives like information. For planning purposes, lots of data is used, but plans are subject to frequent changes. Goals are integrated organizational and personal goals. The integrative functions best in non-hierarchic organizations. Written report communication goes together with elaborate discussion aimed at expressing the interrelationships between reports.

Driver & Mock measured decision style using two main categories of instruments: first, a performance-based approach, second, a self-concept based approach. The self-concept based approach utilizes the CXSD scale for decision style assessment. The CXSD is a (self-description) questionnaire similar to the questionnaire used in the determination of an MBTI score.

Witkin [1950] developed a one-dimensional cognitive style test that measures an individual's field-dependence. Based on test results, a classification into two cognitive styles is made: field-dependent and field-independent. Field-dependent individuals, when compared to field-independent individuals, are not able to differentiate an object from its context. Field-independent individuals are able to break up a basic configuration into smaller parts. This implies that field-independents can be considered to perform better on tasks requiring analytical capacities. So, field-independency is related to high-analytic whereas field-dependency is related to low-analytic types. The distinction between high and low-analytic types has been used in many cognitive style studies [see e.g. Doktor & Hamilton, 1973; Lusk, 1979; Benbasat & Dexter, 1979]. The degree of field-dependency is measured by means of an embedded figures test. In such a test the



subject is asked to find a simple form in a complex figure and to trace it in pencil directly over the lines of the complex figure. In this test, the simple form must be found in exactly the same size and the same proportions and it must face in the same direction within the complex figure as when it appears alone. The simple form can be observed for as long and as often as is convenient for the subject. However, it cannot be observed at the same time as the complex figure. Scores are calculated, depending on the type of embedded figures test,<sup>10</sup> on time required and the number of correct answers.

The distinction between analytic and heuristic cognitive styles [see e.g. Mock *et al.*, 1972; Vasarhelyi, 1977] is very similar to the distinction between high and low analytic cognitive styles. The main differences lie in the measurement method and the fact that a high analytic cognitive style is contrasted to a heuristic style rather than to low-analytic characteristics. As for the measurement method, when classifying an individual with regard to his cognitive style no embedded figures test is used. Instead, a self-concept questionnaire, similar to a verbal protocol analysis, is presented to the subjects. On the basis of subjects' responses the researchers label each individual as being 'heuristic', 'analytic', or 'indeterminable'. A heuristic decision approach is characterized by *ad hoc* sensitivity analysis, trial-and-error, muddling through, and satisficing behavior. A low-analytic cognitive style is characterized by the inability to break up information into smaller, more understandable pieces. In this respect the appropriate analogy is not so much between heuristics and low-analytics as between an heuristic decision approach on the one hand and an intuitive information-acquisition style and feeling information-processing style on the other hand. However, Macintosh [1985] describes the low-analytic cognitive style as (among other things) trial-and-error based. So there is a continuous gradation between low analytics and heuristics.

Tolerance for ambiguity is a personality variable which has been used in several psychological and accounting studies [see e.g. MacDonald, 1970; Gul, 1984.; Pincus, 1990]. Individuals who exhibit a high tolerance for ambiguity perceive no problems when faced with ambiguous situations or materials, they do not need to have the world divided up into strict dichotomies without intermediate options. Information, in this sense, is allowed to be interpretable in a number of different ways. Individuals with a high tolerance for ambiguity excel in the performance of complex tasks. The term 'tolerance for ambiguity' is somewhat misleading since it covers not only a general tendency to tolerate ambiguous situations but also a preference for this kind of situation. Individuals who exhibit a low tolerance for ambiguity experience ambiguous situations as threatening. They want the state of affairs to be put into simple black and white opposites.

MacDonald [1970] reports on various measurement concepts regarding tolerance for ambiguity. One performance-based test measured the extent to which

individuals held on to an original response despite evidence to the contrary. In this test a set of pictures in which a dog was gradually transformed into a cat was presented sequentially. It was assumed that those who maintained their initial response (the picture representing a dog) had a low tolerance for ambiguity. Other tests are self-concept based, presenting subjects with a number of questions (typically 16 to 20) to be answered with 'agree' or 'disagree'. A high score indicates a high tolerance for ambiguity whereas a low score indicates a low tolerance for ambiguity.

A departure from general personality tests is made in management control and management accounting, where leadership styles and performance evaluation styles are investigated<sup>11</sup> [see e.g. Hopwood, 1972; Seiler & Bartlett, 1982; House *et al.*, 1990]. Research reveals that the Hopwood classification<sup>12</sup> serves well as a basis for further research provided that a set of intervening variables<sup>13</sup> is inserted which accounts for the lack of consensus between researchers about the performance effects of different performance evaluation styles [van de Poel, 1986]. The Seiler & Bartlett classification<sup>14</sup> was used to determine individuals' preferences for participative and rigid budgetary systems. It appeared that the three-dimensional classification could predict group membership (participative, rigid systems) with a significant degree of accuracy, but had no significant predicative ability as to the individuals' attitudes towards and perceptions of the budgetary system. House *et al.* [1990] describe a well-known categorization of leadership styles as (1) structuring (2) considerate, and (3) charismatic styles. It was found that the charismatic leadership style can be empirically distinguished from structuring and considerate leadership styles by means of a simple self-concept questionnaire. Seen in relation to subordinate performance, the followers of charismatic leaders reported having goals of higher intrinsic value, higher interest in the task, less role ambiguity, and more self-assurance than the followers of structuring or considerate leaders.

Despite the multitude of cognitive styles, personalities, decision approaches, leadership styles and performance evaluation styles, there is a degree of overlap between all of these.<sup>15</sup> The question as to the 'one and only' optimal classification scheme remains unanswered. However, the Myers-Briggs type indicator has proven to be a fairly good representation of actual cognitive characteristics, since it captures a broad part of an individual's cognitive domain [see e.g. Keen & Bronsema, 1981; Macintosh, 1985]. So it seems an appropriate choice for an application in accounting and auditing.

### 3.3 Capturing human judgment in verbal protocols

An individual's personal characteristics, measured along any scale, must be reflected in that individual's decision behavior. Analysis of (pre-)decisional behavior is conducted as a research methodology to determine the features of

subjective decision processes. Verbal protocol analysis is one of the most widely-applied research methodologies within the broader category of pre-decisional behavior analysis. In appendix D a brief historical sketch of verbal protocol analysis in relation to some mainstream psychological movements is given in order to place the methodological and technical parts of verbal protocol analysis in their broader context.

### **3.3.1 Methods for VPA**

Two methods used for VPA deserve special attention as they can provide the closest reflection of the underlying cognitive processes: concurrent verbal reports and retrospective verbal reports. Both methods are concerned with direct verbalizations of specific cognitive processes.

#### **Concurrent verbalization**

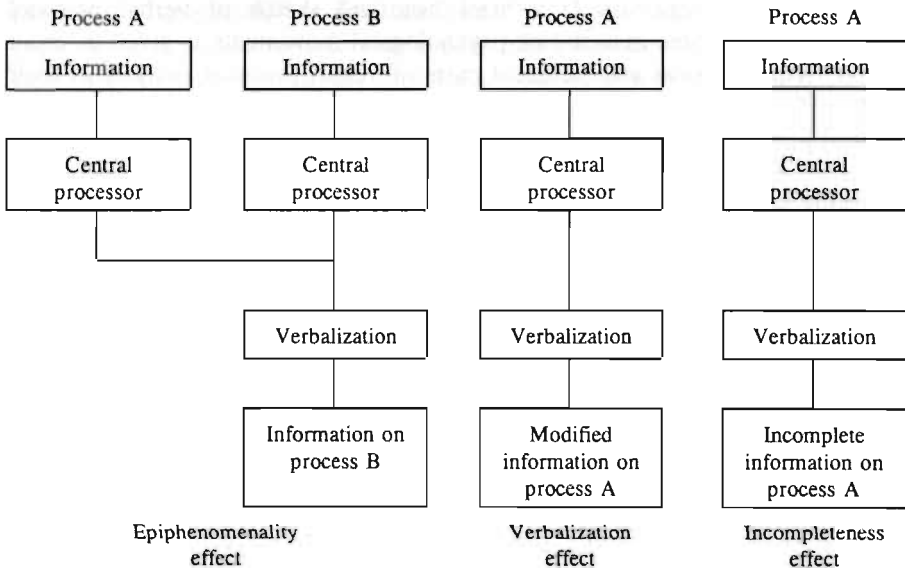
If information is verbalized at the time the subject is attending to it, the procedure is labelled concurrent verbalization. Thus the respondent verbalizes his thoughts at the time the cognitive process under study takes place. This means that when applying concurrent VPA there must be evidence that the verbalizations only report on the actual thought process, not on an activity that occurs in parallel with that process and which is independent of it. Such verbalizations could contain irrelevant information (epiphenomenality argument).

Furthermore there must be evidence that verbalization of the cognitive process does not influence the process (verbalization argument). Finally there must be evidence that thoughts are complete representations of the underlying process (incompleteness argument). These criticisms are summarized in figure 2.

Ericsson & Simon [1984] conducted a literature survey with regard to the potential negative effects of concurrent verbalizations (and of retrospective verbalizations, see below). It appears that neither the epiphenomenality argument, nor the verbalization argument, nor the incompleteness argument can be completely maintained.

#### **Retrospective verbalization**

If information about a cognitive process is verbalized after that cognitive process occurred, the procedure is labeled retrospective verbalization. This means that the cognitive process under study is completed at the time verbalization takes place, and that it cannot then be altered. If that verbalization is concerned with literally recalling the information contents of a problem representation then the retrospective verbalization may be labeled a recall.

**Figure 2: Criticisms of concurrent VPA.**

There are some major problems that must be dealt with when applying retrospective verbalization as a research methodology. For example, a respondent who is asked to report information that was never heeded (that is: taken into STM) could be forced to give an answer that is based on information provided in the question or other information retrievable from LTM instead of just saying: 'I don't know'. Also, the criticisms of concurrent verbalizations as summarized in figure 2 also apply to retrospective verbalization, except that it is the fact of knowing that retrospective verbalizations are to be given, rather than the verbalization itself, which could change the cognitive process under study.

There are two possibilities of accessing LTM: retrieval and inference. Information can be retrieved from LTM, a process which is generally considered onerous and is often unconsciously rejected by humans if it is possible to infer information from LTM. Thus if a chunk of information is not present in LTM, humans often don't know that it is not present because they have tried to infer that information instead of retrieving it. This problem must especially be dealt with when asking 'why' questions in retrospective protocol sessions. For example, the researcher in an auditing experiment might ask: 'Why did you decide on lowering materiality by 10%? The respondent (R) could interpret this question in a number of ways:

1. Why did you decide on *lowering* materiality by 10%?

R would answer: 'Because there are contextual factors such as net income fluctuating in the three years preceding the audit, which urge the auditor to

allow some safety-margin in his materiality assessment'.

2. Why did you decide on lowering *materiality* by 10%?

R would answer: 'Because lowering desired audit risk was not the appropriate measure in the light of surrounding circumstances'.

3. Why did you decide on lowering materiality by 10%?

R would answer: 'Because income having fluctuated in the three years preceding the audit was the only aspect that induced an adjustment of the materiality threshold. If there had been other aspects, such as the company being listed on the stock exchange, then materiality would have been lowered by say 20%.

This illustrates why identical existing knowledge structures could produce different information, dependent on the way a question is asked or a probe is inserted. In the ideal case the retrospective report is therefore given by the respondent immediately after the problem has been solved, because only then will the bulk of information required still be in STM, and only then does the researcher not have to provide the respondent with specific information about what to retrieve. Instead a general instruction could be given, for example: 'Please report everything you can remember about your thoughts when you were solving the case'.

There remains another problem in using retrospective verbalization. When retrieving information from LTM, the respondent could access memory structures other than those created in the just-completed cognitive process. Humans make use of heuristics when making decisions. The heuristic addressed here is 'representativeness' [see e.g. Hogarth, 1980] which can best be described as the phenomenon that humans, when making decisions, compare the characteristics of the input variables of the problem on hand with those of input variables that were used in preceding decision situations. This means that humans who use representativeness to simplify their decision processes construct a categorization of problems which they implicitly refer to when making decisions. The probability of this occurring increases if the respondents have just solved a series of similar problems which they can classify within that categorization. This means that the respondent, when retrieving cognitive structures from LTM, may access information that was heeded some time before, but that is strongly associated with the information heeded during the cognitive process under study. For example, if an auditor acts as a respondent in a VPA study, his experimental judgments could be influenced by his existing knowledge if he has a client that is very similar to the client in the case description for the experiment. Müller [1911] proposed solving this problem by instructing respondents to report only details that they can remember heeding at the time of the experimental session, thus urging them to make a selection which in turn requires a more critical attitude during verbalization.

When retrospective probing is used as a means for gathering information about an

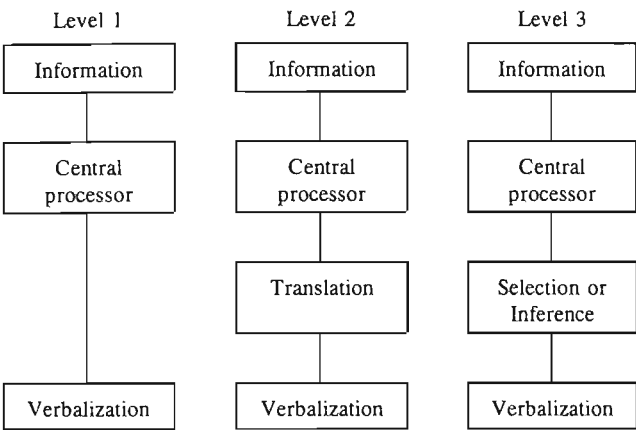
individual's decision processes, the criterion of specificity in reporting needs to be considered when studying the effects of verbalization. It may be useful to distinguish between probing for general information and probing for specific information. When probing for general information, several different ways of arriving at a subjective description of decision processes can be followed. First, subjects are aware of the general procedures they are using and report these directly, second, subjects are aware of some episodes of their decision processes and report a generalization made from these, third, subjects can remember some specific tasks and re-generate the decision processes used in completing these tasks, reporting on the generalizations made in the process, and fourth, subjects can report on a variety of knowledge contents about how a specific task is to be completed. Because it is extremely difficult to determine what kind of reporting behavior subjects are exhibiting, the method of probing for general information cannot be relied upon to retrospectively produce data about actual decision processes. This problem can be easily handled by asking subjects to report on specific instances instead of general procedures. Unfortunately, this method also has some major shortcomings since an individual's memory for applied cognitive processes may be lacking detail and may be incomplete due to an automated, non-reported utilization of certain cognitive processes after a number of repeated applications. However, the degree to which dysfunctional effects of verbalization occur may be dependent on the level of verbalization.

### 3.3.2 Levels of verbalization

Whether the verbalization is concurrent or retrospective, information may be recoded. The degree to which recoding must intervene between the time information was heeded and the time a corresponding verbalization was generated is generally represented in a three-level categorization. Level 1 is when information is verbalized in the same form as it was acquired from the central processor (no recoding). For example, if a rational test person is asked to pronounce the word 'materiality', he can, under normal circumstances, correctly respond. Level 2 verbalization is when a translation process takes place between heeding and verbalization. This is the case if the internal representation of the information is not in verbal code but in a somewhat compact, idiosyncratic nonverbal form. For example, if a respondent is given a description of an organization structure, he may make an internal representation in a nonverbal, graphical or visual imagery form. Verbalization of his thoughts would require additional processing (i.e. translation) in order to make an audience understand what he intends to express. Level 3 verbalization takes place when there is a selection or an inferential process between heeding and verbalization. Selection, for example, would occur if a respondent is asked to report all material errors that he encounters during solving a case: he must test whether the heeded information (error of a certain amount) matches the desired type (above a certain materiality threshold) before reporting it. In figure 3 the three

levels of verbalization are schematically represented. These levels occur within concurrent as well as retrospective verbalization.

Figure 3: Three levels of verbalization.



Ericsson & Simon [1984] argue that thinking aloud in concurrent protocol sessions will not alter the cognitive processes' speed, course and structure, provided that level 1 conditions are met. Under level 2 conditions, speed may be lowered and completeness may not be guaranteed. However, course and structure remain unchanged. In contrast the course and structure of cognitive processes which do not meet level 1 or level 2 conditions, may be changed by having respondents verbalize their thoughts. A distinction in levels of verbalization can be made in simple protocol tasks. Presumably, increasing task complexity makes the assessment of verbalization levels more difficult and thus makes the effects of verbalization on the underlying cognitive processes more indeterminate.

3.3.3 Techniques in VPA

When applying VPA the researcher first chooses among the main research methods: concurrent or retrospective verbalization. Within these categories some specific techniques can be used to objectively elicit thoughts from participating subjects and to transcribe these thoughts into some uniform coding system. Among these techniques are: inserting probes only once, segmenting stimuli into smaller chunks, giving reminders to keep verbalizing during silent periods, reprocessing transcripts in order to identify meaningful statements, and graphical representations.

### Objective elicitation of cognitive structures

During protocol sessions respondents usually do not produce the information required without some probing to stimulate some kind of action. For example, the respondent might be given the general instruction: 'please think aloud while you are solving this case'. When the test person hesitates in his verbalizations or simply forgets to report, the researcher could ask: 'what are you thinking right now?'. This question reduces the probability of interviewer biasing, since it simply urges the test person to verbalize his own thoughts as precisely as possible. In general there are at least three techniques which are used in VPA to increase verbalization by slowing down the underlying cognitive processes:

- a. Inserting a probe only once, so as to minimize the stimulus given to the respondent to access certain memory structures. The respondent thus needs to construct an internal representation of the problem and is not able to rapidly recognize changing stimulus characteristics. For example, in a protocol session regarding materiality the researcher does not give any additional information, such as 'remember that there is no board of outside directors', which could increase the speed of the underlying decision process or could bias outcomes.
- b. Segmenting stimuli into smaller chunks of information which are available on request or which are presented at a slower rate. For example, when a case is presented via a computer screen, information pages can be accessed by means of menu selection so that only small parts of the total information are visible at one time. The menu selection procedure slows down the speed of the cognitive processes under study.
- c. Giving reminders to keep verbalizing during silent periods ('keep talking', 'please, think aloud', 'what are you thinking about now?', etc.).

In retrospective VPA it is hardly possible to stimulate the respondent to increase his verbalizations without increasing the risk of biased outcomes. The best method of conducting retrospective VPA experiments is not to ask any further questions once the respondent has stopped verbalizing for a fixed period of time. The rationale of this can be easily seen when the retrospective process is studied. During the cognitive process the respondent receives stimuli which he first stores in STM. However, as new stimuli emerge the capacity of STM may be exceeded. This leads to a shift in memory from STM via the cognitive system to LTM. So after completion of the cognitive process only a small part of the stimuli which were received are still in STM, while the bulk is in LTM. In retrospective VPA the respondent is asked to verbalize the thoughts that he had during the cognitive process which has just been completed. The respondent needs to dig into STM as well as into LTM to find the relevant thought processes. If the retrieval of memory structures from LTM takes too long the risk that irrelevant memory structures may be addressed increases. However if the researcher fixes the time he will wait for further verbalizations too short, the risk of incompleteness increases. So there must be an optimum time interval between the last verbalization and the moment the



researcher breaks off the retrospective session.

VPA offers ample possibilities to use advanced technical equipment. In early protocol experiments, verbalizations had to be written down or analyzed during the protocol sessions as there were no audio recorders available. The development of the tape recorder, and later the cassette recorder and the video camera, opened new perspectives for VPA with greater numbers of test persons and extensive analyses, leading to greater statistical validity. Automated analyses using computer applications were also initiated [Waterman & Newell, 1971; Bhaskar & Simon, 1977]. An interesting field within VPA, which could be promising in future, is derived from physiological psychology: EEG analysis. An EEG (Electro Encephalogram) is produced by a machine called the electro encephalograph which registers the electrical impulses of the human brain via a set of electrodes connected to the head of the test person. EEG techniques are used mainly in sleep research and in epileptic research. To date, no successful applications of EEG techniques in VPA are known, due to the tremendous technical problems which must be overcome in the development of EEG machines. If these problems are solved EEG techniques could produce the most objective data on cognitive processes and thus be the most powerful instrument within VPA.

### **Uniform coding systems**

Once the protocols are registered on tape, they can be transcribed for analysis. Several coding systems have been developed for transcription. They can be classified into two main categories: verbal systems and graphical systems. VPA usually involves categorizing verbalizations based on the process criterion, that is, according to the process that generated a specific verbalization. It has not proved possible to devise a classification for protocol analysis which is applicable in every situation (comparable to the well known taxonomy of businesses as used in auditing) as language is so elaborate and oral expressions have a tremendous richness. However, it is possible to give some general guidelines to make a classification unambiguous and unique.

Waterman & Newell [1971] attempted to automate protocol analysis because they wanted to increase objectivity. In doing so they implicitly developed a set of rules for the representation of verbalizations. They restricted the scope of their research to the following constraints:

- a. the system was primarily for their own use,
- b. the system would be developed for a specific task,
- c. the system would be constructed by means of prototyping,
- d. only a limited number of researchers would be involved,
- e. the aim was full automatization and not some man-machine symbiotic system.

The first restriction is especially relevant for the discussion of coding systems since

each coding system in VPA is tailored to the needs of the researcher. However, some common features which form the basis for all verbal coding systems can be defined. Within a well-defined task two structures are to be distinguished: first, the problem space, which specifies the kinds of knowledge a respondent can have about that task, and second, the production system, which consists of an ordered set of condition-action elements. Each knowledge state, as a part of the problem space, is connected to another knowledge state by means of a production system. There is an analogy with static and dynamic human information processing research: a description of problem space elements is purely static whereas a description of production systems is purely dynamic. A verbal coding aimed at representing decision processes should naturally encompass both structures.

VPA transcripts can be encoded at a very detailed, task-dependent level (i.e. low level encoding). However, if the aim is to find some general characteristics of human decision processes, a more aggregate task-independent approach must be followed (high level encoding). Newell & Simon [1972] conducted a VPA experiment based on a cryptarithmic task.<sup>16</sup> A typical low-level encoding was applied in order to map subjects' decision processes during problem-solving. The possible knowledge states and production systems were exclusively defined, so that the coding would be as uniform as possible. However, the applicability of the results is therefore limited to theories concerning human behavior during cryptarithmic problem-solving, rather than to the general characteristics of human behavior. Similarly, Bhaskar & Simon [1977] tried to develop a symbiotic automated protocol analysis system for solving thermodynamic fluid problems. They made stronger assumptions as to the number of coding categories, so that a low level encoding was effected.

At the other extreme are high level coding schemes which define general categories, make less and weaker assumptions and hence are applicable over a broad range of tasks. Humans differ widely in their knowledge content and knowledge utilization in different situations. A possible point of criticism of the aggregation of protocol data into a high level encoding scheme could be the loss of information. However, the trade-off between the generalizability and accuracy of a theory must be made, searching for some kind of optimum. For audit judgment research a combination of high level and low level encoding is preferable from a validation point of view. This will be discussed further in the next chapter. Model-based coding can be considered a form of low level encoding since only information relevant to the task and model is identified. No attempt is made to handle all the verbalizations in the protocol. Ambiguous material will thus be removed from the analysis or classified in a residual category labelled: 'irrelevant' or 'comments'.

Although there are as many VPA coding schemes as there are VPA studies [see

e.g. Ericsson & Simon, 1984], some general remarks about the encoding process can be made. Before starting an encoding, the vocabulary of respondents must be adapted to a uniform vocabulary. For example, if a subject consistently uses the word *tolerance* instead of *materiality*, a search for the word *materiality* would not reveal encodable verbalizations. Therefore all the appearances of the word *tolerance* must be translated into the word *materiality*. Provided that the translation is conducted properly, without judging synonymity too lightly, no significant negative effects on reliability are to be expected.

After encoding the vocabulary, the transcript will be scrutinized to find statements which cannot be understood without adding explanatory terms such as explicit goal specifications. For example: if a respondent utters the following phrase: 'looking at balance sheet total and gross turnover I would consider the minimum of these', the following goal specification would be added by the encoder: 'in order to assess materiality' (or any other, more symbolic form). This encoding step produces an adjusted protocol which can easily be differentiated into meaningful statements.

The statements will be aggregated to a greater or less degree, depending on the level of encoding desired. Aggregation, in this context, implies a segmentation of the protocol into instances of general processes. In this type of encoding, verbalizations are classified into segments which are constructed to test certain hypotheses. For example, if the aim is to test hypotheses about relationships between certain personality variables and problem-solving behavior, it might be convenient to encode the protocols into information search activities, calculation activities, theoretical fact reporting, deduction, retrieval from memory, and cognitive process reporting, leaving the unclassifiable items to the category of comments.<sup>17</sup> Short-term memory contains inputs and outputs to processes, not the processes itself. This implies that a person verbalizing his thoughts only reports on these inputs and outputs. Verbalizations about decision processes stem from inferences made by the respondents and may not be reliable. In order to deal with this kind of potential unreliability, or just to find evidence for non-reported data on decision processes, a process inference will be made by the encoder. The classification above is an example of such a coding. The encoding into a sequence of instances of a (limited) number of general processes postulated by the model does not take account of the sequential relations among processes. Instead, only the relative frequencies of processes are determined, in order to test hypotheses. A variant of this approach is appealing: the relatively unstructured verbal protocols could be summarised in a vector of process frequencies which could be analyzed by standard statistical methods. However this technique requires highly skilled and precise encoders who base their encodings on a detailed coding manual. Increasing inter-coder consensus should increase reliability, and in fact no more objective measurement instruments exist.<sup>18</sup>

Coding manuals must give instructions as to the criteria the encoder should apply when segmenting protocols. When coding is model-based, with only a limited number of categories distinguished, uniform segmenting is especially important since the impact of between-segment shifts is relatively high. Criteria vary on a continuum from objectively determinable to subjectively determinable. For example, pauses, intonations and time-intervals are relatively objective when compared with content-based criteria such as explicit line-end markers, separating ideas and concept-counts or word-counts. Furthermore, subjective criteria are further sub-divided, by Waterman & Newell [1971], into lexical and topical representations. The former consisting of uttered words, phrases and sentences, and the latter of single task contents. As will be seen in chapter 6, this distinction leads to a completely different analysis of verbal protocols.

If the purpose of VPA is to describe decision process sequences, a graphical representation can be constructed as a supplement to a verbal coding scheme. Newell & Simon [1972] developed the problem behavior graph technique. This technique is based on the distinction between knowledge states and production systems which was mentioned above. Each transition between states is marked by means of an operator. Protocol analysis should reveal in what knowledge state a certain decision is taken and what operators are used in doing so. Other graphical techniques include the decision tree approach, in which a decision process is broken down into testable sub-goals which are set out in a tree-diagram, and the transition net in which a sequence of actions is represented by means of arrows and boxes similar to the problem behavior graph.

### **3.4 Expert modelling as a method for studying audit judgments**

Building expert systems is a means of learning more about how particular judgments are made. Steinbart [1987] developed an expert system ('Auditplanner') for audit planning stage materiality judgments for just this reason. He stipulates that expert systems explicitly contain information on the role played by various pieces of information and on the reason for using that information. He uses the expert system construction process as a way to explicate domain-specific knowledge in terms of if-then rules and their relative importance in the decision to be made.

#### **3.4.1 Determination of expertise**

In building expert systems it is important to separate real expertise from plain knowledge. Bedard [1989] stipulates that: 'at present, there exists no generally accepted definition or measure of expertise'. Yet three elements that should be included in any definition of expertise have evolved [see Bonner & Lewis, 1990]: education, experience and innate personal characteristics or 'ability'. Furthermore,

expertise is domain dependent. So expertise can be defined as the knowledge content, within a specific domain, which has evolved as the result of education and experience within the framework of a person's intuition, ability and talent. This definition would intimate that experts will be difficult to distinguish since the subjective elements of intuition, ability and talent are difficult to assess. However the population from which the selection must be made can easily be reduced to manageable proportions by using the objective elements as the first selection criteria. Education is measured by the examinations an individual passes, and experience is measured by the number of years an individual has been working in a certain problem environment. For example, when searching for auditing experts, the required educational level is the official CPA exam (and subsequent registration in the CPA register), and the required number of years auditing experience is defined by years of employment by an auditing firm. The more complex problem of finding talented auditors with a certain instinct for their profession can be handled by searching for auditors who have had a very successful career within an accounting firm [see e.g. Choo, 1989]. By following this selection method a subjective criterion is turned into a more or less objective (or intersubjective) criterion since consensus among top decision-makers in an auditing firm about the career developments of certain individuals should be an acceptable, and presumably the only, substitute for objective quality measurement regarding a person's professional performance. However, the method is very general and will not necessarily reveal information about task-specific expertise. Choo [1989] lists some other, more direct, criteria for expertise measurement which are also cited most often in the literature: self-rating and independent rating. However, here again we are dealing with surrogates for real expertise which are flawed in some respect. Self-rating is affected by the opinions of those non-experts who consider themselves experts, and independent rating, for example by supervisors, is affected by biases - due to raters' lack of understanding about the nature of expertise - and perceptual errors. These remarks suggest that a subjective element is always present in expert selection.

Expert systems can be developed on the basis of the knowledge of one or more experts. If only one expert's knowledge is elicited and put into the system the risk of excessively low system-reliability increases. On the other hand, if a multitude of experts are used for knowledge elicitation, the risk of not achieving consensus on certain items may increase. The latter could ultimately lead to a significant quality reduction (e.g. results that are too general or too difficult to interpret) as a result of compromising. If there is considerable consensus within a specific knowledge domain then multiple-expert system development is not impossible. This has some important implications for human information processing research, since consensus studies in a wide range of problem areas have on the whole been conducted only to assess decision accuracy. In relation to expert system development, a second justification can be added. For example, research revealing that there is consensus

among auditors with regard to certain audit decisions implies that multiple-auditor expert systems can be built, provided that other prerequisites for expert system development are met, thus increasing system reliability.

### 3.4.2 Knowledge elicitation techniques

Having determined the experts who will provide the knowledge to develop the expert system, the knowledge elicitation method to apply must be selected. Neale [1988] reviews a number of knowledge elicitation methods, arguing that most research on expert system development has given only limited attention to a thorough documentation of these methods. Each method has its advantages and disadvantages in relation to the domain, the task, the expert(s) and the knowledge engineer. The following main categories<sup>19</sup> of knowledge elicitation methods will be discussed: (1) interviews, (2) observation, (3) multidimensional techniques, and (4) verbal protocol analysis.

#### Interviews

The simplest form of knowledge elicitation is interviewing experts. A large subset of interview structures, aimed at eliciting knowledge as reliably as possible, has been developed. However, all interviewing strategies suffer from the difficulty that experts may present knowledge which substantially diverges from that which they actually use in practice. Furthermore, the interviewer may introduce biases through the way questions are asked.

The structured interview is characterized by a great number of questions which are listed in a pre-designed questionnaire. It is possible to skip questions, within a fixed framework. This implies that the interviewer needs to understand the method behind the questionnaire as well as the terminology of the domain as it is used by the expert. The formulation of questions is a crucial factor in structured interviewing; care must be taken that the expert's knowledge is accessed in the most natural way.

In a tutorial interview the expert is asked to prepare an introductory talk, outlining the main themes and ideas of the knowledge domain. Concepts are then extracted from an integral transcription of the talk. This method is especially suitable when there are no textbooks or other written reference materials in the field. The teachback interview has some resemblance to the tutorial interview. However, in the teachback interview the expert gives a personal introductory talk to the interviewer followed by a re-presentation by the interviewer to the expert. If the expert agrees with the teachback of the interviewer, they share the same concept and the interviewer has understood the expert.

Focused interviewing is a relatively unstructured form of knowledge elicitation

since only the topic is prepared in advance by the interviewer, not the exact questions to ask. This method is naturally more flexible than other interviewing methods. This technique is extremely suitable for explorative purposes.

Interviews based on the explicit goal presentations of the knowledge engineer, and subsequent questioning as to the evidence necessary to distinguish that goal from other alternatives, produces some kind of decision-tree model of an individual's knowledge. A second decision-tree based approach is found in the classification-reclassification methodology. Working from facts to goals and answering questions regarding which goals are supported by which facts, and then reclassifying the goals into sets of facts supporting these goals, the expert finally arrives at facts that are directly observable by the interviewer. The division of knowledge domains into observable facts which ultimately, via several subgoals, lead to the final goal is closely related to this methodology. Finally, the laddered grid approach is suitable when the knowledge domain is hierarchical. The higher the branch on the 'ladder', the higher the level of abstraction and the smaller the number of defined characteristics. In downward laddering the expert is typically seeded with a concept and asked to provide examples of that concept together with a set of criteria which distinguish each of them. In upward laddering the expert typically is seeded with a concept and asked to provide the higher level concept of which the seeded concept is a part.

If it is possible to construct exhaustive lists of all the possible faults and symptoms in a knowledge domain, then a systematic symptom-to-fault link procedure can be followed for knowledge elicitation. Such a list is presented to the expert in order to find relationships between symptoms and faults. In this approach the interviewer may gather data on the intermediate reasoning steps in order to find evidence about the expert's reasons for going to the next step in the symptom-to-fault link.

The ethnographic interview is a means of accessing an expert's knowledge through friendly conversation from which it is not clear (to the expert) that the intention is to elicit knowledge by means of frequently asking for explanations of uttered statements.

Introspection and forward scenario simulation are interviewing methodologies in which the expert gives a general description of how he would solve a certain problem or a certain case. Data is obtained via thinking aloud processes.

Retrospective case description and critical incident description are both retrospectively oriented methodologies. In the former the expert is asked to describe how he handled one or more typical cases. In the latter the expert is asked to describe his experiences when handling remarkable or difficult cases.

As will be seen in the section on protocol analysis as a knowledge elicitation methodology, there are many overlaps between introspection, retrospective case description, critical incident description and forward scenario simulation on the one hand and verbal protocol analysis on the other hand.

### **Observation**

Instead of asking the expert about his decision processes, he can be observed during decision-making. One important restriction to this technique is that the knowledge engineer must be almost an expert himself, since he receives only limited verbal cues from the subject. Two observational methodologies, differing as to the level of verbal input, can be distinguished: non-participative dialogues and participant observation.

Dialogues between experts and knowledge engineers during actual task performance can be recorded on video or audio tapes. These dialogues are transcribed literally and analyzed as in verbal protocol analysis. However, in some knowledge domains there is a preponderance of knowledge which cannot be articulated verbally and hence must be observed. In that case a participant observation technique must be used, implying that the knowledge engineer must spend some time as a practitioner in the problem area which is being investigated.

### **Multidimensional techniques**

In multidimensional techniques, concepts are related to each other via dimensions that are to be identified by the expert. The output of such analyses is often quantitative and easy to analyze statistically.

Concept card sorting according to dimensions chosen by the expert, and subsequent analysis, reveals a limited number of relevant higher-level concepts based on a much more extensive set of lower-level concepts. However, when the underlying dimensions are not known another methodology must be followed: multidimensional scaling (MDS). When applying MDS a comparison is typically made between each combination of two concepts in order to find underlying similarities. Multivariate statistics play a very important role in this analysis. Since multidimensional scaling techniques are based on what are intended to be exhaustive lists of concepts, dependency on verbal reporting by experts is reduced to a minimum. For this reason MDS is appealing. However, a much-cited shortcoming of MDS is that detail is (partially) missing, reducing the technique to no more than a discussion piece on which to base further investigations. Proximity analysis complements MDS, since it is also based on a list of concepts but additional information about local structuring is retained. The concepts are to be compared to each other and sorted into three piles: related, may be related, and not related. The related concepts are then rated on a 0-100 scale indicating the relative strength of the relationship when compared to other pairs of concepts.



The repertory grid methodology, like MDS and related techniques, is based on what is intended to be an exhaustive list of concepts. The expert's perception of a certain problem is represented in a matrix consisting of elements (rows) and constructs (columns). The constructs are to be identified by the experts by determining in what (bipolar) dimension each pair or triad of elements differ. The elements which are not used to determine the dimension of difference are rated by the expert, usually on an interval scale from 1-5. After completing the repertory grid, each column (dimension or construct) consists of ratings on the chosen scale for each element in the rows. Multivariate statistical analysis is then applied to find individuals' underlying concepts, resulting in a dimension-based clustering.

### **Verbal protocol analysis and its application in expert system development**

As has been discussed earlier, VPA produces information on knowledge contents which suffers in only a limited way from problems arising from the effects of verbalization: incompleteness, epiphenomenality and idiosyncrasy. However, there are two arguments that might lead to a rejection of VPA as a methodology for knowledge elicitation for expert system development. Firstly, the subjectivity in encoding argument. However, all other knowledge elicitation methodologies discussed suffer from this kind of problem, and there are techniques - which are described in this chapter - that mitigate this argument. Secondly, an observation must be made concerning the use of VPA in knowledge elicitation from experts, as opposed to novices: experts generally have trouble with verbalizing their thoughts at a sufficiently basic level to meet the formal requirements for information which is to be stored in the knowledge base of an expert system [Waterman, 1986]. They often omit in their verbalization those chunks of information which are obvious to them. To deal with this shortcoming of VPA in eliciting experts' knowledge, a series of protocol sessions is usually followed by one or more refinement stages in which experts comment on the protocol-based prototype. Waterman [1986] mentions three techniques for extracting knowledge from domain experts, which in fact are methods of improving the system's performance: system refinement, system examination, and system validation. In *system refinement*, the expert whose knowledge has been put into the system provides a number of problems he has encountered in practice. These are processed by means of the prototype expert system, and the expert compares the system's output with his own judgment. In *system examination*, the same expert who was used in building the system criticizes the rules/frames that are incorporated into the system. Finally, in *system validation*, the cases used to build and refine the system are presented to outside experts for a last refinement round.

Experts are generally busy people. This implies that any knowledge elicitation process must be as efficient as possible. Clearly VPA is very time-consuming and therefore does not meet a possible goal of limiting the experts' involvement. However, if VPA yields better representations of thought processes with less bias,

this is an argument in itself against using less reliable methods for extracting data from experts. If experts are willing to invest a considerable amount of time in the construction of an expert system, VPA supplemented by system refinement techniques is therefore recommended above any other knowledge elicitation technique. VPA consists of a large subset of techniques. It is noteworthy that the interviewing techniques of introspection, retrospective case description, critical incident description and forward scenario stimulation, as well as the observational technique of non-participative dialogue analysis, can all be considered as elements of VPA if VPA is defined as 'the set of techniques used to extract, record and analyze an individual's knowledge in such a way that negative interviewer and respondent effects are reduced to a minimum by having respondents talk and think aloud during problem-solving'. There is no argument against this definition [see e.g. Payne, Braunstein & Carroll, 1978; Weber, 1980; Biggs, Mock & Watkins, 1988; Johnson, Jamal & Berryman, 1989; Williams, 1990].

### **Knowledge representation**

The choice of modelling media, or more generally the restrictions imposed by the hardware and software which are to be used to model the elicited expert knowledge for the further development of an expert system, could be considered in the stage of knowledge acquisition. However, some methodologists argue that a simultaneous consideration of the knowledge model and the system model is highly inefficient when trying to cope with a construct as complex as expertise [Hickman *et al.*, 1989; Steels, 1992]. For that reason, expert knowledge is described on three levels: the knowledge level, the symbolic level, and the physical level. If one wants to gain an insight into general concepts, tasks, and mental models that are used during cognitive processing, a description at the knowledge level is most appropriate. On the symbolic level, knowledge is described in terms of a language that can be directly entered in the knowledge base of an expert system. On the physical level, which is the least abstract level, knowledge is described in a form that can be directly entered in a computer system. This description is close to the machine code and hence is subject to hardware and software restrictions.

### **3.5 Synthesis**

This study must follow two lines of research, the first based on a research question about how auditors make materiality decisions in planning an audit, and the second based on questions about how human information processing research, and audit judgment research in particular, might be conducted. This chapter has outlined some technical aspects of the latter facet of this study. Justifications for choosing the three research approaches discussed in this chapter can be found in the literature on human information processing [see Libby, 1981], since a dynamic approach to human information processing provides more direct information on how humans make decisions than a static approach. As has been demonstrated,

personality assessment in relation to VPA and process tracing are research methodologies which can be considered dynamic when compared to lens modelling and studies of probabilistic judgment. Furthermore, expert system development is a well-established means of learning more about how humans make decisions. This study attempts to combine VPA and personality assessment sessions and the knowledge elicitation sessions in order to get a comprehensive picture about how auditors make judgments in general and in materiality and risk decision making. In accordance with the proposed general audit judgment model, the focus is on the three main elements of dynamic audit judgment studies: psychological profiles as measured by personality questionnaires, demographic characteristics, and auditors' problem-solving behavior. The main function of the current chapter is to clarify some concepts and ideas from within the fields of personality assessment, VPA and process tracing, and expert system development, in order to make the empirical analyses which follow in chapters 5, 6, and 7 more understandable.

## Notes

1. One research methodology which should make it possible to evaluate the validity of cognitive style questionnaires would be introspection. However, due to the subjectivity inherent to this methodology, it has hardly been applied in accounting.
2. As will be demonstrated later, the computer analogy has been a starting point for much research into human information processing.
3. Cf. the multivariate clustering technique in statistics, which also identifies class membership relations.
4. This notion is close to practice, as one would imagine, since the auditor's opinion formulating process more often than not involves negotiations between auditor and the client's management.
5. Perceiving is the process by which people become aware of things and other people around them. Information acquisition is thus clearly part of the perceiving dimension.
6. Judging entails the process of coming to conclusions about what has been perceived. Information processing is thus clearly part of the judgment dimension.
7. For those who think that art and accounting are absolutely incompatible, an interesting, albeit somewhat artificial, book is on the market: *Art & Accounting*, by B.S. Yamey.
8. The number of questions differs between applications.
9. However, the analogy between the pairs agreeable-true and multiple solution-one solution is not as close as the analogy between perception style in the Jungian typology and information utilization in the Driver & Mock decision style model.
10. Three embedded figures tests have been developed in particular: the original embedded figures test which must be scored individually, the children's embedded figures test which must also be scored individually, and the group embedded figures test which can be scored in a large group.
11. Personality refers to attitudes and beliefs whereas style refers to the methods by which individuals make decisions.
12. Hopwood [1972] distinguished three performance evaluation styles: (1) a budget-constrained style, (2) a non-accounting style, and (3) a profit-conscious style.
13. For example: budget participation (high/low) and environment (uncertainty/certainty) have been used as intervening variables.
14. Seiler & Bartlett [1982] distinguish three personalities: (1) authoritarian, (2) independent, and (3) flexible.
15. Similarly, a high correlation should be observed between the results of the tests as described in this section.

16. The problem was to assign digits to letters in such a way that the vertically calculated sum of DONALD+GERALD equalled ROBERT, given that D equalled 5. During problem-solving the subjects were to talk aloud. The encoding of the resulting transcripts was heavily dependent on the task. For example, the operator elements were: process column for information about a certain letter, assign a value to that letter, generate the possible values of that letter, and test whether the assigned digit is valid for that letter. As can be seen, this kind of coding applies only to a specific cryptarithmic task.

17. This classification is highly effective for the research hypotheses which will be discussed in the following chapters. However the framework of criteria used to construct this classification is subjectively determined, following the tradition of tailor-made coding schemes in verbal protocol analysis.

18. The same reasoning was found in the determination of audit decision accuracy.

19. As is the case in almost every classification, the line of demarcation between categories is sometimes difficult to draw.



---

## Research methodology

---

### 4.1 Introduction

This chapter describes the overall research methodology as applied in chapters 5, 6, and 7. Section 4.2 discusses the research design. Section 4.3 describes the task which subjects were asked to perform. Section 4.4 discusses the measurement of information access sequence and attention duration. Section 4.5 describes the development of a technique for transforming verbal protocols into data, and presents summary data. Section 4.6 presents a set of research questions focusing on the relationship between auditors' demographic characteristics and psychological profiles, and their problem-solving behavior.

### 4.2 Research design

A two-round audit judgment research project was set up to describe auditor knowledge acquisition and decision processes while making audit planning stage materiality and risk judgments. The first round consisted mainly of a series of VPA and automated process tracing sessions, as well as personality and demographic questionnaires. The second round consisted mainly of a series of expert system refinement and examination sessions, which were aimed at gathering data about how auditors make planning stage materiality and risk decisions. The verbal protocols and the automated process traces were also analyzed in this round.

#### 4.2.1 Subjects

All data was gathered from one group of 25 subjects drawn from 4 large auditing firms in the Netherlands, in two strongly interrelated interview sessions with each subject. Subjects were selected by top executives in each audit firm. The criteria for selection were as follows:

- The subjects should have at least 3 years of auditing experience and must have reached at least a senior level within the firm.
- The subjects selected within each firm must have a client population that is heterogeneous with regard to firm type (production, trade, financial, service, non-profit, agricultural, etc.).
- The geographical dispersion of the offices from which they work must be great.

#### 4.2.2 Procedures

In the first round of sessions, a case (see section 4.3) was presented to the subjects via a computer screen. They were asked to think aloud while trying to solve the case. Two recordings were made: first, the computerized information that was accessed by subjects during problem-solving was traced by means of a resident time-stamping program, and second, subjects' verbalizations were registered on audio and video tape and transcribed literally (concurrent protocols). When the subjects had solved the case they were asked to fill out two personality questionnaires: a tolerance for ambiguity test and an information acquisition/processing test. Their years of audit experience, firm affiliation, and educational background were also recorded. Research question 1 was formulated as follows: 'Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?'. At this stage the data required to answer this research question could be analyzed. Subjects were then asked to think aloud while trying to remember what their thoughts were while solving the case. These verbalizations were also recorded on audio and video tape and transcribed literally (retrospective protocols).

After the protocol and process tracing sessions were finished, the research continued in round two in two separate directions.

Firstly, the concurrent and retrospective protocols, together with audit manual and textbook knowledge, were used to construct a prototype expert system. This prototype expert system was presented to 17 subjects for system refinement purposes. When the system refinement was complete, the refined prototype was presented to the remaining 5 subjects<sup>1</sup> for system examination purposes. Finally, the knowledge which had been entered into the knowledge base of the prototype expert system was analyzed and graphically represented by means of diagrams showing the task structure, semantic network, and methods (see chapter 7). The data on materiality determining factors was then used to answer research question 3: 'What factors determine audit planning stage materiality thresholds?'.

Secondly, the literally transcribed verbal protocols were coded into relevant statements and subsequently coded into operators and task categories (see chapter 6). Also, data from the automated process trace was analyzed to obtain evidence about auditors' information access sequence and information attention duration. The protocol data, process trace data, personality data and demographic data were then used to answer research question 2: 'Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?'.

Figure 1 represents the stages in the research project. The procedures which are



directly linked to the research questions of this study have been marked by means of the labels Q1 to Q3.

Concurrent and retrospective VPA, and more direct interviewing techniques, have been combined in this study. The combination of concurrent protocol sessions with either subsequent sessions which are intended to erase the experts' short term memory, followed by a recall session, or with retrospective protocol sessions, is not unconventional [Ericsson & Simon, 1984; Boshuizen, 1989]. Also, expert system refinement is often combined with expert system examination in order to extract knowledge from an expert by discussing practical or hypothetical cases, guided by the prototype expert system [Waterman, 1985]. There are several techniques for knowledge elicitation, which may be used separately, but they provide more reliable knowledge descriptions when applied in combination [see e.g. Neale, 1988].

#### 4.3 Task design

When applying process tracing and VPA, a case must be constructed which enables the researcher to gather domain knowledge as objectively as possible, that is without interfering in the problem-solving process. This is the most important limitation of VPA as a knowledge elicitation technique: the case is the bottle-neck. Therefore, a great deal of attention must be paid to the case construction and validation.

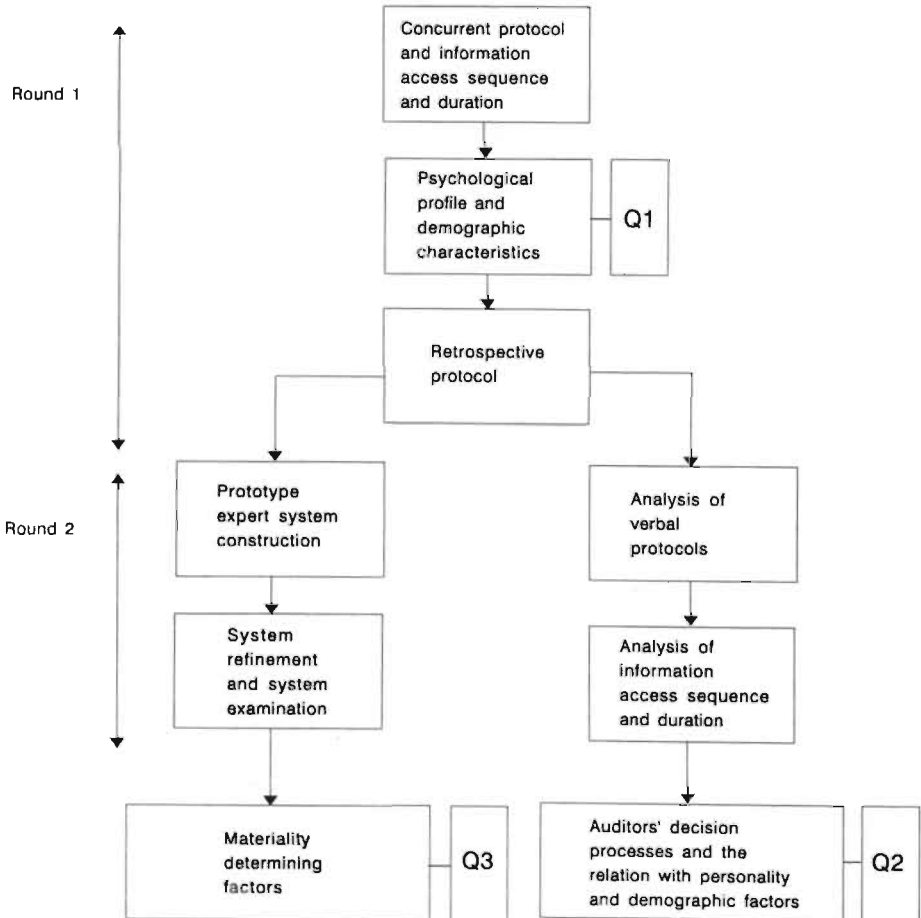
A draft version of the case was drawn up, based on theoretical knowledge and practical experience. This was then pre-tested by two experienced auditors at a medium sized auditing firm (7 and 12 years auditing experience respectively, both auditors being at the manager level) in order to refine it and to detect how the expert organizes his domain knowledge, represents concepts and hypotheses and handles inconsistent, inaccurate, or imprecise elements in the case description. The pre-test auditors introduced some new terms (e.g. different concepts of materiality) and gave comments on the format of financial data as presented via the computer screen and on the sequence in which information was displayed. Finally, they were asked to judge whether or not the case was appropriate for eliciting the factors determining materiality and whether or not it was a realistic problem setting (assuming the proposed modifications were implemented).

The pretest auditors thought the case might be *inappropriate* in terms of cue contents (materiality determining factors). How this subjective feedback was to be used presented a problem. VPA is aimed at eliciting knowledge (in this research: materiality and risk determining factors) by means of think-aloud techniques without any interviewer interference, yet this subjective element is required in the development stage of the research, for the case must be related to domain

expertise. Yet, if the process which is intended to elicit domain expertise is based too closely on the expertise of the pretest auditors its objectivity would be questionable: the research would assume the form of a conversation. This dilemma can be solved by adding a number of relevant, irrelevant or doubtful information items which should prompt the subjects to comment in terms of agreeing or disagreeing.

In round 1 of the research project, subjects were presented with the computerized case description. In appendix H, a summary of that case is made. The case consisted of two introductory pages, a problem definition, and 26 information pages which were freely accessible on menu selection by striking one of the

Figure 1: Procedures in the audit judgment research project.



number keys. The case described a firm which was involved in a takeover, and required an auditor's opinion on the financial statements. The problem which the subjects were asked to solve read as follows:

1. How do you evaluate the materiality of the various items in the balance sheet and the profit and loss account, in other words, what factors are relevant in passing judgment on the reliability of the annual accounts?
2. What quantitative standards of materiality would you use in this audit?
3. What additional information would you like to have had access to, when certifying the annual accounts?

This setting and problem definition was purposely rather vague, with some uncertain elements which the subjects would have to deal with. For example, the materiality problem could have been defined as: 'what quantitative materiality threshold do you use to plan the nature, extent and timing of the audit procedures?'. Such a problem definition might induce subjects to calculate the materiality threshold on the basis of a materiality table as incorporated in their audit manual, without thinking - and hence verbalizing - about underlying concepts and qualitative factors. As will be explained in chapter 6, the degree to which subjects are tolerant to ambiguous situations, like the way in which the problem has been defined in the case description, may be linked to their treatment of the case.

Note that in the problem definition no reference has been made to the concept of audit risk. Yet this concept came up while subjects were thinking aloud. In chapter 7, a distinction is made between qualitative factors which might influence materiality and hence are to be considered true materiality determining factors and qualitative factors which might influence audit risk, but are incorrectly considered to influence materiality. Because many subjects confused the concepts of materiality and audit risk, the analyses in this thesis must necessarily focus on materiality as well as on audit risk.

At the beginning of each session in round one, the research project was briefly described to the subject. This comprised an explication of the goal of the project, a brief explanation of audit judgment research in general and of verbal protocol analysis and personality research as applied in the current research, and the follow-up.

Before each concurrent protocol session the following standard text was presented orally to the subjects:

'Please try to solve the following case, thinking aloud as much as possible. Remember, you are the expert and the opinion of the researchers is of no importance. Therefore, we will not answer any questions as to the contents of the information pages you are going to see. If you don't understand a certain piece of information, please report as clearly as possible why you don't understand it. The only questions we are willing to answer are those regarding the operation of the computer and the program that handles the case presentation. You may start now.'

After each subject had indicated that he had reached a conclusion on the case, he was asked to fill out two forms: a compressed Myers-Briggs type indicator questionnaire as adapted by van de Poel [1986], and a Tolerance for Ambiguity Questionnaire (MAT-20) as adapted by MacDonald [1970]. Additional demographic data was gathered, using a debriefing questionnaire, on the subjects' years of audit experience, firm affiliation, and formal education. In appendix G, the demographic and psychological data are summarized. Auditors' personal characteristics are expressed by six variables. The two variables that constitute psychological profile are: Myers-Briggs type indicator scores (MBTI; to be divided into information acquisition and information processing scores), and tolerance for ambiguity scores (TFAM). The three demographic variables are: number of years of auditing experience (NYE), firm membership (FIRM) and education (EDU). After completing the concurrent VPA session, the respondents were given the tolerance for ambiguity and Myers-Briggs type indicator questionnaires (see section 5.2). These questionnaires served two purposes: first, information on the respondents' psychological profiles was obtained which could be correlated to performance during the concurrent VPA session [Vaassen & Hayes, 1990; Vaassen, Hayes & Baker, 1992], and second, short term memory regarding the case was erased, so that the respondents had to 'dig' into their long term memory when retrospectively verbalizing their thoughts.

Finally, after answering the questionnaires the first round of sessions was concluded by orally presenting the following standard text to the respondents [cf. Ericsson & Simon, 1984, p. 149]:

'Please state everything that you remember you were thinking about while you were working on the case.'

In answering this question, subjects were considered to give an indication of what elements of their problem-solving processes they recalled. This information was used for the construction of the prototype expert system and hence for the analysis

of materiality determining factors (see chapter 7).

In round 2 of the research project, a prototype materiality expert system was constructed and refined by means of cases which were brought in by 17 of the 25 subjects who participated in round 1 of the research project. In this round, each subject was asked to provide at least one case which he had encountered in practice and which was to be solved by means of the prototype expert system.<sup>2</sup> Each time the expert system asked a question, the subjects could insert the keyword 'why' in order to have the system explain why that specific question was asked. By recording and analyzing the comments that were given by the subjects the prototype expert system was refined and additional data on materiality determining factors came available. Also in round 2, the refined prototype expert system was presented to 5 of the remaining 8 subjects, to examine its performance and gather more data on materiality determining factors. For these sessions two structured cases were developed which contained all the information that was necessary for having the prototype expert system determine a materiality threshold (see appendix L).

Ultimately, a set of materiality determining factors was established, based on the literal concurrent verbal protocol transcripts, audit textbook and audit manual knowledge, literal retrospective protocol transcripts, and comments received in the system refinement and system examination sessions. The approach that has been followed in determining this set of materiality determining factors is aimed at acquiring as many potential explanatory variables as possible. So the description of materiality determining factors which results is exploratory. This approach is highly pragmatic: further research into the weights, the correctness and the completeness of these factors must be conducted in order to validate the set of materiality determining factors. However, since auditing is a pragmatic profession, a summary of materiality determining factors is presented anyway (chapter 7).

#### **4.4 Measurement of information access sequence and attention duration**

When solving the computerized case, each subject was free to access any information page by menu selection and to have that specific information page active. Measuring the information access sequence is more complicated than measuring information attention duration.

##### **4.4.1 Information access sequence**

The information access strategy could be sequential or directed. A sequential strategy would imply a top-down selection conforming to the sequence in which menu items were presented, whereas a directed strategy would imply a search for the information considered relevant to the problem, thus accessing a limited number of information pages. The top-down sequence, constituting the standard

ranking of information pages, is given in appendix I.

Because the subjects were free to select any information page via menus, some items were not selected by all the subjects, whereas some other items were selected more than once by some subjects. This resulted in two problems in comparing the information access sequences of the different auditors and in comparing each auditor with the standard access sequence:

1. What ranking should be used for sequences in which the subject did not select certain information pages? Here the question is the rank of non-accessed pages in the analyzable sequence.
2. How to analyze sequences in which a certain information page was accessed more than once by the same subject? Here the question is to the rank of repeatedly accessed pages in the analyzable sequence.

Since each information page was accessed by at least one subject (total attention duration for each page > 0), the maximum rank in the access sequence would appear to be 29. However the first three pages, consisting of two introductory pages and the problem definition, were programmed to appear on the screen regardless of the subject's preferences. When determining correlation measures for information access sequence, therefore, these pages must be removed from the analysis. This leaves 26 pages and a corresponding maximum rank of 26.

There are two alternatives in dealing with the first problem: (1) delete cases with missing values (i.e. missing rankings due to non-access) 'listwise', that is, an information page is removed from the analysis entirely if any subject did not access it and (2) assign a maximum rank of 27 to the pages that were not accessed by a subject. It was found that only two pages (page 4 and page 9) were accessed by all subjects. The first alternative is therefore not feasible since only two information pages would be left in the analysis (pages 4 and 9). On average, 32% of the pages were not accessed. This leaves the second alternative. Although the artificial ranking of 27 is arbitrary, this solution to the missing value problem is appealing, since all cases will be incorporated in the analysis. As the maximum possible ranking within the set of existing values is 26, the nearest ranking for the non-existent values is 27.<sup>3</sup>

The second problem, concerning what ranking to assign where there was repeated accessing of the same information page, is solved by taking the rank of the first access of each page. Again, this approach is highly arbitrary. However, there are no arguments for taking any higher-ranked access for the analysis. The average ranking for a given page might be used but, considering the function of repeated accessing (to reread relevant information) the average ranking per page was seen as an indicator of information complexity rather than of access sequence.

#### 4.4.2 Information attention duration

When a page was accessed more than once, the attention duration used in the analysis was the sum of attention durations for that page. The attention duration measure is thus an indicator of information relevance. Here the first three pages, which were programmed to appear on the screen regardless of the subject's preferences, were also incorporated in the computation of correlation measures.

#### 4.5 Verbal protocols as data

Ericsson & Simon [1984] mention several issues which researchers must deal with if they want to use verbal protocols as data in behavioral experiments. Two of the main issues is the suitability of subjects' verbalizations as scientific data, in the first place, and how it is to be processed to transform it into data that can be statistically analyzed.

##### 4.5.1 The development of a technique for transforming verbal protocols into data

In VPA, *operators* are concepts indicating a verbalized action or activity which is part of a subject's decision process. Based on Newell & Simon's [1972] theory of human problem-solving, verbalizations can be encoded as goals, operators and knowledge states. These are the three dimensions of the problem space. Since subjects' decision *processes* are being studied, the *operators* are of primary concern [Biggs & Mock, 1983]. A classification of operators into four general categories can be made [Mock *et al.*, 1993]:

1. task structuring operators (set goals and subgoals),
2. information acquisition operators (information search, direct reading, algebraic calculation, information retrieval, assumptions),
3. analytical or inferential operators (conditional judgments, evaluations, decision support, generate queries),
4. action or choice operators (generate alternatives, audit decisions, decision rules, other decisions).

This classification is tailored for audit research purposes [see e.g. Biggs & Mock, 1983; Biggs, Mock & Watkins, 1988; Mock *et al.*, 1993]. If the number of encoding categories were increased, the number of arbitrary allocations would also increase since the borderline between each pair of categories would then become less clear. However it might be convenient to develop more specific encoding categories to match a specific set of research hypotheses. In this dissertation, a more customized encoding system was in fact developed, taking the Mock *et al.* classification as a starting point.

Verbal protocol encoding is a highly subjective process. Therefore, the first problem was how to increase the objectivity of encoding. Two alternatives have

been proposed in the VPA literature [e.g. Wouters & De Jong, 1982; Biggs & Mock, 1983; Vermunt *et al.*, 1986; Mock *et al.*, 1993]: the multi-encoder approach and the single encoder/elapsed time approach. Typically, a small number of persons (2-4) have been used to fulfil encoding tasks. Biggs & Mock [1983] applied a multi-encoder approach where each concurrent protocol was independently encoded by two persons. Mock *et al.* [1993] applied both a single encoder approach, with each protocol being encoded twice, with a significant (three to six months) time interval between, and a multiple encoder approach, with each protocol being encoded by two encoders.

In this study, encoding took place in two separate stages. In stage one, each literal protocol transcription was reduced to a set of relevant statements. In stage two, the relevant statements were encoded into operators and tasks.

Relevant statements are verbalizations, each consisting of one full sentence which has meaning on its own. Five encoders were asked to transcribe (all or parts of) the literal protocols into relevant statements. A sixth encoder checked the encodings and reconciled these with the encoders. The instructions presented in table 1 were given to the encoders.

Table 1: Instructions for transcribing protocols into relevant statements.

- 
- |   |   |
|---|---|
| 1 | Search for key words such as 'if' and 'then'.   |
| 2 | Record only one information element per statement.  |
| 3 | Don't record physical actions which the subject undertakes (he starts with...).   |
| 4 | Check strange sentence constructions or sentences containing strange words by re-transcribing the original tape recordings.   |
| 5 | Try to record standard wordings as much as possible in order to increase inter-subject comparability.   |
| 6 | It is not the word usage (synonyms), but the content that is relevant.  |
| 7 | Each statement must contain information which is relevant to the case in its own right.   |
| 8 | Take your time to summarize parts of the original transcription into one elementary information element containing only the essence (however, do not alter the sequence of verbalizations). |
-



The reason for applying a two-stage encoding process is that some persons have great linguistic abilities whereas some others have not [Breuker *et al.* 1986]. Subjects who have great linguistic abilities would be expected to produce protocols that contain well-formulated statements that are both efficient and effective in communicating messages. On the other hand, persons who have less linguistic abilities will produce protocols that show hesitations and have no smooth narrative style. Since linguistic abilities are not indicative of superior performance in auditing tasks, these linguistic differences must be mitigated. This is accomplished by transcribing the literal protocols into relevant statements. Moreover, by doing so, an indicator of efficiency of communication can be developed. This indicator, labeled 'relevance ratio', is also discussed in this chapter.

Vermunt *et al.* [1986] apply a two-stage protocol analysis methodology which is similar to the approach followed in this dissertation. They first make a segmentation of literal protocol transcriptions into relevant statements, which are defined as meaningful entities that contain exactly one process description. Breuker *et al.* [1986] mention a number of factors that will affect thinking aloud. Among these are the verbal abilities of the subject. The resulting differences relate to verbal intelligence. They also outline a methodology for preparing protocols that contain only relevant information.

In stage two of the protocol encoding the relevant statements were categorized into operator categories and task encoding categories. This was in order to obtain vectors of category frequencies for statistical analyses. In table 2, the operator categories and operator definitions which were used are presented. The encoding scheme as discussed here is appealing because it is customized to the research questions as asked in this study, and because the process of audit decision-making - as represented by the general audit judgment model - can easily be recognized.

The experimental case consisted of an investigation and an audit section. In this chapter a set of research questions regarding this type of encoding and personality characteristics is developed. This explains why a secondary encoding scheme, consisting of a classification of verbalizations into 'audit', 'investigation', 'both but determined' or 'both but indeterminate' is applied. This division is represented in table 3.

Under the heading 'operators', four categories can be distinguished: information search, information retrieval, judgment, and decision. In order to code the relevant statements into operator categories, a distinction is first made between compound statements (i.e. statements that consist of, or can be rewritten into, an if-part and a then-part) and singular statements (i.e. statements that have a single information content). The compound statements are coded based on table 4. The singular statements are coded based on detailed descriptions of operator categories.

**Table 2:** Relevant statement operator encoding categories.

Operator	Description	Code
Information search	Specific request for information.	IS
Information retrieval	Knowledge which has previously been stored in memory is retrieved from memory.	IR
Judgment	A pre-decisional evaluation or judgment is made about audit procedures to perform, about the direction of the audit, about the collection of evidence, etc.	J
Decision	A decision about the planning stage materiality threshold is made.	D

**Table 3:** Relevant statement task encoding categories.

Task	Description	Code
Audit	An explicit verbalization regarding the audit part of the case.	A
Takeover price verification	An explicit verbalization regarding the takeover price verification part of the case.	T
Determined	Both audit and takeover price verification are implicitly or explicitly referred to.	DT
Indeterminate	Neither audit nor takeover price verification are implicitly or explicitly referred to.	I

**Table 4:** Encoding of compound statements.

THEN	IF	
	Goal specification (‘want to’ or ‘not want to’)	Observable phenomenon (‘is’ or ‘is not’)
Fact generation	n/a	IR
Audit procedure	IS	J
Materiality decision	D	

A goal specification is made when the subject utters a statement like: ‘If I want to verify the takeover price then ...’. An observable phenomenon is a fact that may be true or false. E.g.: ‘If you are unfamiliar with accounts receivable, then ...’. A fact generation is made when the subjects verbalizes a statement like: ‘... then

inventory turnover is high'. A verbalization of an audit procedure which will be performed is an indicator of information which is needed for decision making. E.g.: '... then the auditor should look very carefully to such a balance sheet account.'. Finally, a materiality decision is made when the subject verbalizes a statement like: '... then the planning stage materiality threshold is 250,000).

### *Information search*

An information search operator is assigned when the if-part of a compound statement contains a goal specification. For example:

*If the goal is to determine a quantitative materiality threshold, then I will start with taking note of the P/L and the B/S.*

Or:

*In order to make things clear, the auditor should ask management: 'if there are misrepresentations in the annual accounts, to what amount would you hold me responsible for that?'*

It is difficult to verbalize a statement that contains both a goal specification and a fact generation: a goal specification can only be combined with an audit procedure or a materiality decision. Auditing is concerned with evidence collection and the comparison of collected evidence with norms. A more general description of a statement relevant to the auditing process would be 'search for information' (*IS*). This is exactly how encoding took place.

If the relevant statement is singular, then an *IS* operator is assigned when a specification is made of information that is needed for decision making. For example:

*Also of relevance when considering the account 'machines' is whether or not the terms of depreciation are determined realistically.*

Or:

*The auditor wants to know whether special categories of buyers are involved.*

In summary, an *IS* operator is assigned when specific information is asked for because the auditor considers this information relevant for decision making.

### *Information retrieval*

An information retrieval operator is assigned when the if-part of a compound statement contains an observable phenomenon (fact that can be true or false) and when the then-part of that statement contains a fact generation. For example:

*If the bank has negotiated certain profitability requirements in the finance contracts, then the auditee has an incentive to keep balance sheet amounts within certain boundaries.*

When verbalizing an *IR* operator, the subject retrieves existing knowledge from memory. This can easily be seen when considering the above examples: the auditor has certain knowledge in his memory (if X then Y) and during the process of decision-making he applies that knowledge in order to solve a problem. However, knowledge application can also be verbalized by means of a singular statement. In that case an *IR* operator is assigned when a fact is generated or when an expectation is expressed. For example:

*The gross profit indicates the margin between purchase and selling.*

Or:

*An increase in equity may be a result of reservation of profits or of a share issue.*

In summary, an *IR* operator is assigned when existing knowledge is retrieved from memory and verbalized.

### *Judgment*

A judgment operator is assigned when the if-part of a compound statement contains an observable phenomenon (fact that can be true or false) and when the then-part of that statement contains an audit procedure (in a broad sense, including calculation, classification, or summarization). For example:

*If the annual accounts are reconciled with the general ledger, and the valuation of the accounts is checked, then I would not spend very much time on auditing the internal control system.*

Or:

*For fixed assets which have been valued on the basis of historical costs, it is important to determine whether the valuation on the basis of current costs deviates strongly from this.*

If a *J* operator is verbalized by means of a singular statement, then it can be detected by trying to answer one of the following questions:

- Is evidence being evaluated in order to make a materiality decision?
- What audit procedures are performed?
- Is the audit mainly system oriented or mainly data oriented?

For example:

*The value of equity in a takeover assignment is more cautiously determined than in an audit assignment.*

However most *J* operators will be in the format of if-then statements.

### *Decision*

A decision operator is assigned when the then-part of a compound statement contains a materiality decision.

*According to an accounting firm guideline, a turnover of 34 million leads to a materiality threshold of 250,000.*

Like *J* operators, most *D* operators are in the format of if-then statements. However, singular statements may also occur. For example:

*The audit planning stage materiality threshold is 250,000.*

In summary, judgments are pre-decisional verbalizations. Each materiality decision is preceded by at least one judgment. Generally, 'judgments' will outnumber 'decisions'. Since the task was to determine a quantitative materiality threshold, only statements containing verbalizations of that quantitative materiality threshold are coded '*D*' for 'decision'. The statements that contain a decision other than materiality (concerning for example the audit program) are coded '*J*' for 'judgment'. This indicates that each subject's protocol will contain not more than one *D* operator. If no *D* operator is assigned, then the subject could not make or forgot to make a final decision on the materiality threshold.

There is a relationship between these coding categories and the classification of knowledge in memory as either declarative or procedural knowledge. Declarative knowledge is knowledge that consists of *facts and data* which are stored in memory. Procedural knowledge consists of the *skills and process* knowledge required to perform a certain task. Since auditors are mainly involved in cognitive tasks, procedural knowledge in auditing refers to cognitive skills. As has been explained in chapter 3, the difference between declarative and procedural knowledge can best be compared to the difference between respectively 'knowing

that ...' and 'knowing how ...'.

There is also a relationship between the afore-mentioned coding categories and the audit judgment model as presented in chapter 1. Through repeated judgment and decision-making, a person learns from experience. Statements that refer to these kinds of activities are coded 'J' or 'D'. Learning from experience is a necessary condition for gaining expertise. This type of knowledge is stored as procedural knowledge in memory. By learning facts from, for example, textbooks, a person stores declarative knowledge in memory. If a person retrieves procedural and declarative knowledge from his memory, then an *IR* operator is assigned. If a person verbalizes a statement that contains a request for additional information - in other words: a request for data that the subject considers relevant for decision making - then an *IS* operator is assigned. Statements coded '*IR*' are more general than statements coded '*J*'. This is because *IR* operators refer to declarative or procedural knowledge which is stored in memory and which is therefore not specific to the case on hand. *J* operators are much more specific to the case. Here another difference between *IR* and *J* operators is observed, since a judgment (*J*) is the combined result of an information search operation (*IS*) and an information retrieval operation (*IR*). So if there is any indication that information which is not in memory is combined with existing knowledge in order to arrive at a judgment, then a *J* operator is assigned.

The encoding into task categories, as represented in table 3, is much simpler to accomplish. Any *IS*, *IR*, *J* or *D* operator may be coded '*A*' for 'audit', '*T*' for 'Takeover price verification', '*DT*' for implicitly or explicitly determined 'audit' or 'investigation' or '*I*' for 'both audit and investigation' (i.e. indeterminate). If there is evidence that nothing other than the audit is considered when a subject utters a statement, then it is coded '*A*'. For example:

*The audit planning stage materiality threshold is 250,000.*

If there is evidence that nothing other than the takeover price verification is considered when a subject utters a statement, then it is coded '*T*'. For example:

*When determining the takeover price, the magnitude of the result is much more important than the constituting elements of that result.*

If the subject makes it clear that he needs to distinguish between the audit and the takeover price verification part of the case, then the relevant statement is coded '*DT*'. For example:

*Materiality in a takeover investigation is something completely different to materiality in a financial audit.*

If nothing in the relevant statement indicates whether the audit or the investigation is considered, then it is coded 'T'. For example:

*The usual categorization within the balance sheet is important to gain an insight into the liquidity.*

#### 4.5.2 Relevant statement, operator and task encodings

Appendix J presents an example of an encoding into relevant statements. The encoding was conducted by five persons (one research assistant and four faculty members) with a basic theoretical audit background, but without any practical experience in auditing, and one person (a faculty member) with two years of practical auditing experience. One encoder (a research assistant) performed an integral encoding of all verbalizations by all the subjects into relevant statements,<sup>4</sup> with the other four inexperienced faculty members performing only a partial encoding of 6 pages of literal protocol transcript each (randomly selected). Based on a pairwise comparison of relevant statements (maximum two encoders for each set of statements), there was a sample agreement of about 52% of the relevant statements. Because this sample agreement was rather low, the integral encoding was checked and discussed by the experienced faculty member in order to increase the robustness of the relevant statement encodings.<sup>5</sup>

The encoding of relevant statements into operator categories was conducted by three persons. One research assistant (coder 1) without any practical experience or coding experience, a 'registered accountant' (coder 2) with two years of practical auditing experience and some coding experience (audit judgment courses), and a faculty member (coder 3) without any auditing experience but with some coding experience (a three-day audit judgment course) integrally and independently performed the encoding task. Reconciliation took place in two stages. In the first, coder 1 and coder 2 reconciled their encodings. This revealed a reconciled set of operator frequencies. The Kappa coefficient between these two encoders was 0.38 (raw agreement = 0.63). Because of this low agreement the third encoder independently re-coded the statements. The reconciled encodings of coder 1 and 2 were then reconciled with the encodings of coder 3. The Kappa coefficient was then 0.58 (raw agreement = 0.76) which was considered acceptable when compared to previous research findings [Mock *et al.*, 1993: raw agreement = 0.66; Biggs & Mock, 1983: Kappa = 0.66, raw agreement = 0.75].

The encoding of relevant statements into task categories was conducted by two

persons (coder 1 and coder 2). Since the Kappa coefficient was 0.82 (raw agreement = 0.89) a second round of encoding was not considered necessary.

As an example of the encoding into operators and tasks, the partial protocol transcription from appendix J can be encoded as shown in table 5. The reconciled operator and task encodings and the Kappa coefficients are summarized in table 6.

Table 5: Example of an encoding into operator categories.

Operator encoding	Task encoding	Relevant statement
IS	I	The balance sheet is relevant for obtaining insight into balance sheet proportions, business activities, and financial structure.
IS	I	The profit and loss accounts of the preceding years indicate the growth rate of the business.
IR	DT	The broad role of the advisor in a takeover and the aspects that he focuses on have nothing to do with materiality.
D	A	According to an accounting firm guideline, a turnover of 34 million leads to a materiality threshold of 250,000.

The basic nature of the protocols that were obtained is shown in table 7. The concurrent protocol sessions generated from 14 to 55 relevant statements, from 201 to 1401 words in these relevant statements, and from 1340 to 6704 words in the literal concurrent protocol transcripts. The ratio between the number of pages accessed and the freely accessible information pages (*COMP*) ranged from 0.42 to 0.92. The time needed to solve the case ranged from 22 to 85 minutes (1344 to 5088 seconds). The retrospective protocol sessions generated from 1 to 16 relevant statements, from 6 to 360 words in these relevant statements and from 134 to 1960 words in the literal retrospective protocol transcripts.

The relevance ratio - as an indicator of efficiency of verbalization - is calculated as the word count of the relevant statements divided by the word count of the literal concurrent protocol transcripts (*WCRS/WCPT*).

4.6 Research questions

4.6.1 Psychological profiles and demographic characteristics

When trying to answer the general research question:



*Q1*

*Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?*

three research questions regarding the uniformity of distribution of personality types among auditors were addressed:

*Q1.1a*

*Do sensing (S) types and intuition (N) types occur equally frequently among auditors?*

*Q1.1b*

*Do feeling (F) types and thinking (T) types occur equally frequently among auditors?*

*Q1.1c*

*Do low and high tolerance for ambiguity types occur equally frequently among auditors?*

If results indicate that there is a preponderance of specific personality types among auditors, then testing hypotheses regarding auditor personality differences in relation to actual problem-solving behavior will not provide significant results.

Two questionnaires were presented to the subjects after the concurrent protocol sessions. In appendix E, the simplified Myers-Briggs type indicator items used in this study are presented. The wording of each item was customized by Van de Poel [1986] to reduce linguistic effects, because not every item in the English original from the Myers-Briggs type indicator lends itself to adequate translation into Dutch. The number of test items was also reduced to 16. Since we are dealing with Dutch-speaking auditors, and there was limited time available for filling out the questionnaires, this version of the Myers-Briggs type indicator has been used in the study. In appendix F, the tolerance for ambiguity items used in this study are presented (adapted from MacDonald [1970]). Both types of questionnaires are well tested. However under specific circumstances (e.g. when applied among a population of experienced auditors) some items could exhibit a low item-total correlation. For that reason a biserial correlation was calculated for all the items in the questionnaires (see appendix E and appendix F). None of the items showed a significant negative correlation with the total score on each test, so all items were retained in the analyses.

**Table 6:** Reconciled operator and task encodings.

Sn	Kappa	Operator frequencies				Kappa	Task codes			
		IS	IR	J	D		A	T	DT	I
1	0.70 <sup>6</sup>	2	24	20	0	0.89	9	13	1	23
2	0.83 <sup>6</sup>	0	19	9	1	0.71	4	8	2	15
3	0.52	2	22	10	1	0.86	5	5	0	25
4	0.34	4	32	19	0	0.84	12	4	7	32
5	0.88	2	21	9	1	0.71	7	10	8	8
6	0.77	14	28	9	0	0.84	8	9	1	33
7	0.82	1	15	12	1	0.85	11	8	2	8
8	0.58	5	20	14	1	0.91	10	6	1	23
9	0.49	9	12	25	1	0.85 <sup>6</sup>	14	6	0	27
10	0.58	1	49	4	1	0.96	17	1	0	37
11	0.35	4	31	16	0	0.68	26	4	4	17
12	0.64 <sup>6</sup>	1	7	5	1	0.86 <sup>6</sup>	6	2	1	5
13	0.75	4	11	3	1	0.67	4	1	4	10
14	0.67	2	21	12	0	0.87	8	4	5	18
15	0.48	6	17	13	0	0.86	4	7	4	21
16	0.57	0	30	13	1	0.64	9	5	3	27
17	0.51	2	16	11	0	0.75	1	10	6	12
18	0.52	4	12	8	0	0.96 <sup>6</sup>	1	3	1	19
19	0.55	4	9	17	1	0.92	2	3	3	23
20	0.47	0	9	15	0	0.76	3	3	3	15
21	0.33	2	20	10	0	1.00	0	5	4	23
22	0.72	1	8	9	0	0.78 <sup>6</sup>	5	3	3	7
23	0.73	1	12	22	0	0.92	3	4	1	27
24	0.44	1	15	20	1	0.69	3	3	4	27
25	0.45	10	26	10	0	0.76	3	11	4	28
Total	0.58	82	486	315	12	0.82	175	138	72	510

Table 7: Overall summary of protocol and information access data.

Sn	Concurrent				Total time in seconds (TT)	Retrospective		
	Number of relevant statements (NRS)	Word count relevant statements (WCRS)	Word count literal protocol transcripts (WCPT)	Complete- ness of infor- mation access (COMP)		Number of relevant statements (NRS)	Word count relevant statements (WCRS)	Word count literal protocol transcripts (WCPT)
1	46	971	3821	0.77	3429	3	56	305
2	29	547	2724	0.62	2027	1	23	382
3	35	685	5196	0.81	1998	5	127	551
4	55	1401	5650	0.69	3406	1	28	196
5 <sup>7</sup>	33	571	2904	0.27	1966	24	473	5296
6	51	1113	3515	0.77	5088	8	170	678
7	29	547	1977	0.77	2802	5	107	587
8	40	815	4883	0.69	2584	16	360	1295
9	47	957	3554	0.81	4215	1	14	174
10	55	1004	6704	0.54	3784	1	6	706
11	51	1165	6230	0.62	4454	2	32	134
12	14	201	1340	0.42	2468	13	230	1736
13	19	429	2255	0.81	3962	1	21	189
14	35	814	5430	0.81	4701	2	28	1089
15	36	732	4783	0.92	4258	12	247	1403
16	44	892	3445	0.62	1759	5	84	366
17	29	659	5601	0.92	2974	14	302	1960
18	24	543	4582	0.54	1344	1	27	289
19	31	720	2014	0.73	2096	9	259	338
20	24	566	2796	0.73	2218	1	30	657
21	32	787	3318	0.58	2513	3	71	298
22	18	527	3219	0.69	1907	2	58	243
23	35	825	4170	0.77	3757	1	24	193
24	37	1045	6537	0.50	3247	2	51	612
25	46	987	6543	0.69	3623	7	107	772
Avg.	36	789	4179	0.70	3109	5	103	631

**Table 8:** Summary of demographic data.

Sn	Information acquisition ( <i>MBTI-S</i> )	Information processing ( <i>MBTI-F</i> )	Tolerance for ambiguity ( <i>TFAM</i> )	Firm ( <i>FIRM</i> )	Number of years auditing experience ( <i>NYE</i> )	Formal education ( <i>EDU</i> ; 0 = NIVRA, 1 = academic)
1	0	0	-2	2	10	1
2	0	0	-4	2	23	1
3	4	6	0	3	27	1
4	2	-2	-6	3	15	1
5	0	2	4	2	24	0
6	2	0	4	2	17	0
7	-4	-2	0	2	17	1
8	2	2	12	1	13	1
9	-2	-2	-2	2	20	0
10	-2	0	2	1	27	0
11	-2	2	0	1	36	0
12	4	4	0	2	12	0
13	0	-2	2	2	12	0
14	-4	2	14	1	12	1
15	-4	-4	2	3	23	0
16	0	4	4	1	25	0
17	0	0	6	1	26	1
18	0	0	-4	3	39	0
19	2	-4	8	4	12	0
20	-2	2	0	4	8	1
21	2	2	8	3	39	0
22	0	4	-2	3	21	0
23	0	-2	6	3	30	0
24	-2	-2	6	3	28	0
25	2	0	10	1	19	0

Only two dimensions of the Myers-Briggs type indicator have been studied: information acquisition and information processing. These are strongly related to

the research questions that are posed in this dissertation, which deal with auditors' decision processes and the information acquisition preceding these processes. The variables to be analyzed are constructed from the subjects' answers to eight questions for each dimension. In order to calculate the variables (one for each dimension) for each respondent, we counted the number of times they did not answer in accordance with the characteristics of a certain type (sensing for the acquisition dimension, feeling for the processing dimension) and subtracted this from the number of times the test person did answer in accordance with those characteristics. This produced a score ranging from -8 to +8 for each dimension. The scores were arranged in 3 groups (score < 0, score = 0 and score > 0). Subjects scoring zero are not categorized as sensing or intuition types on the one hand or into feeling or thinking types on the other hand.

The tolerance for ambiguity test consisted of 20 items that could be answered with either true or false. The same transformation was performed as with the Myers-Briggs type indicator, resulting here in a variable ranging from -20 to +20, with a high score indicating a high tolerance for ambiguity. These scores were also arranged in 3 groups (score < 0, score = 0 and score > 0).

The demographic variables that were considered in this study were: firm affiliation, number of years of auditing experience, and formal education. Table 8 summarizes the demographics and personality test scores for the subjects in this study.

The relationship between demographic variables and personality types must be assessed in order to determine what variables interact when explaining actual problem-solving behavior. As has been discussed in chapter 3, certain personality types may be closely linked to certain educational and professional choices. In order to optimize performance on the job, human resource departments within accounting firms might make personnel selections which reflect firm culture. In some cases personnel selections might (partially) depend on personality questionnaires. There might be differences between firms due to differences in the personnel selection techniques and criteria applied, differences in the homogeneity of organizational cultures, or differences in the degree of structure of firms' audit approaches. Doubts about the reliability of personality assessment methods would be supported if the research revealed a positive or negative correlation between experience and personality, indicating that personality as measured by personality questionnaires is subject to change over time. The research questions identified here are:

#### *Q1.2a*

*Is experience related to information acquisition style?*

*Q1.2b**Is experience related to information processing style?**Q1.2c**Is experience related to tolerance for ambiguity?**Q1.3a**Are there differences in the information acquisition styles of auditors with a NIVRA and an academic (university) education?**Q1.3b**Are there differences in the information processing styles of auditors with a NIVRA and an academic education?**Q1.3c**Are there differences between the tolerance for ambiguity of auditors with a NIVRA and an academic education?*

Earlier research reveals that the audit approaches of the firms in our study have varying degrees of structure [Kinney, 1986; Cushing & Loebbecke, 1986; Bamber & Snowball, 1988; Morris & Nichols, 1988; Bamber & Tubbs, 1989; Bamber et al., 1993]: firm 1 has the least structured approach whereas firm 2 has the most structured approach, and firm 3 occupies a position somewhere in between. Structure may be related to psychological profiles. Hence, the research questions identified here are:

*Q1.4a**Are there differences between accounting firms' as to the information acquisition styles of their affiliates?**Q1.4b**Are there differences between accounting firms as to the information processing styles of their affiliates?**Q1.4c**Are there differences between accounting firms as to the tolerance for ambiguity of their affiliates?*

#### **4.6.2 Psychological profiles, demographic characteristics and problem-solving behavior**

When answering the general research question:

*Q2*

*Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?*

a set of research questions regarding the constituent elements of expertise - in accordance with the general audit judgment model - is developed. By repeatedly making audit decisions, the auditor becomes experienced. Together with audit firm influences (colleagues, audit firm structure and culture, country, client portfolio, etc.), a basic formal education (the educational program that qualifies a person to issue an opinion on financial statements), and certain personality characteristics, 'on-the-job' experience contributes to a certain level of audit judgment quality.

Experts use strategies which are mainly goal-oriented, whereas novices put great emphasis on the immediate consequences of their actions [Anzai, 1987]. Murphy [1990] indicates that experts generally have a more elaborate episodic memory than novices, whereas the two groups have identical semantic memories. The more years of auditing experience a subject has (in combination with a suitable educational background and personality structure), the greater the chance that he has experienced and correctly stored into long term memory a particular problem. Hence, experts see similarities between the current problem and a problem which they have previously solved. In terms of memory and knowledge components: experts' long-term episodic memories contain production systems, categories and schemata that together form templates which the expert uses and modifies to solve new problems [Waller & Felix, 1984]. These templates are not available to novices but, provided that they possess the required procedural and declarative knowledge (as part of their semantic memory), they may be able to solve a problem anyway. However, the underlying problem-solving process will be much slower because they need to make more intermediate steps regarding information retrieval from their memories. This does not mean that experts' problem-solving processes take less time than novices'. Experts may detect many more sub-problems than novices would, causing the total time required to solve a case to increase [Weber, 1978; Colbert, 1989]. These findings from previous research would indicate that there must be a relationship between the number of years auditing experience and the numbers of operators in each of the *IS*, *IR*, *J* and *D* categories. Bedard [1989] stipulates that expertise can be measured as a function of the number of years auditing experience. The general audit judgment model as presented in the first chapter of this dissertation considers expertise development to be influenced by: experience, education, audit firm affiliation, and psychological profile. The experience component in the model leads us to ask:

*Q2.1a*

*Is there a relationship between the number of years auditing experience and the relative number of operators in each category IS, IR, J and D?*

The education component in the model leads us to ask:

*Q2.1b*

*Is there a relationship between the educational background and the relative number of operators in each category IS, IR, J and D?*

The audit firm component in the model leads us to ask:

*Q2.1c*

*Is there a relationship between firm affiliation and the relative number of operators in each category IS, IR, J and D?*

The incorporation of a personality component in the model implies that auditors' psychological profiles might influence their problem-solving behavior. In this study, a psychological profile is considered to consist of the following elements: cognitive style for information acquisition, cognitive style for information processing, and tolerance for ambiguity. Based on the personality component in the model, four research questions are addressed.

As has been discussed in chapter 3, sensing types require more information for decision-making. Therefore the analysis of these subjects' verbal protocols is expected to reveal a preponderance of verbalizations regarding information search activities. The research question here is:

*Q2.1d*

*Do sensing types verbalize more information search operators than intuitive types?*

An irrelevant verbalization is defined as one that either replicates a previous verbalization or that does not contain information which contributes to a solution to the problem described in the case. As has been discussed in chapter 3, feeling types tend to ramble and repeat themselves. This implies that they may need more words to communicate a certain concept than thinking types. We thus expect a positive answer. For each subject, the relevance ratio is calculated as the total number of words in the relevant statements divided by the total number of words in the literal protocol transcriptions. Furthermore, following logical processes of reasoning would imply that more time is needed for decision-making. Thinking types follow these logical processes of reasoning more often than feeling types and hence might need more time for decision-making. The research question here is:

*Q2.1e*

*Do thinking types verbalize more relevant statements and IR operators and do they take more time for decision-making than feeling types?*



High tolerance for ambiguity types may not find it necessary to make judgments or decisions as frequently as low tolerance for ambiguity types would. This proposition would be supported if we find that those who have been typed as having a high tolerance for ambiguity provide relatively more information search and information retrieval operators, whereas those with low tolerance for ambiguity provide relatively more judgment and decision operators. The research question here is:

*Q2.1f1*

*Do those with low tolerance for ambiguity verbalize more J and D operators than those with a high tolerance for ambiguity, and do those with high tolerance for ambiguity verbalize more IS and IR operators than those with low tolerance for ambiguity?*

High tolerance for ambiguity types do not make their choice for one or other task in an engagement explicit, thus accepting an ambiguous situation. On the other hand, low tolerance for ambiguity types are expected to make their choice for the investigation or the general audit explicit in the early stages of their problem-solving process. The research question here is:

*Q2.1f2*

*Do those with a high tolerance for ambiguity verbalize more operators from which it is unclear whether or not the investigation or the audit task is being addressed, as compared to those with a low tolerance for ambiguity?*

Furthermore, experienced subjects exhibit superior performance in recall tasks [Weber, 1980; Boshuizen, 1989; Christ, 1993]. In this dissertation, the information which is accessed is chosen by each subject. When examining subjects' recall, a correction will be required for differences in the information subjects have accessed. The research questions here are:

*Q2.2a*

*Is there a relationship between the number of years of auditing experience and the number of recalled relevant statements, after correction for the amount of information accessed?*

*Q2.2b*

*Is there a relationship between the educational background and the number of recalled relevant statements, after correction for information access?*

*Q2.2c*

*Is there a relationship between firm affiliation and the number of recalled relevant statements, after correction for information access?*

*Q2.2d*

*Is there a relationship between information acquisition styles and the number of recalled relevant statements, after correction for information access?*

*Q2.2e*

*Is there a relationship between information processing styles and the number of recalled relevant statements, after correction for information access?*

*Q2.2f*

*Is there a relationship between tolerance for ambiguity and the number of recalled relevant statements, after correction for information access?*

We measured the between subject agreement on information access sequence and information attention duration in order to assess whether or not there is any relationship between these agreement measures, demographic characteristics and psychological profiles. As for the information access sequence and the information attention duration, six research questions were designed, prompted by the idea that audit manuals must be structured as efficiently as possible. Preferred information access sequences and attention durations could provide guidelines for this structuring. The research questions are:

*Q2.3a*

*Is there a relationship between the number of years auditing experience and the degree of agreement on information access sequence and information attention duration?*

*Q2.3b*

*Is there a relationship between the educational background and the degree of agreement on information access sequence and information attention duration?*

*Q2.3c*

*Is there a relationship between firm affiliation and the degree of agreement on information access sequence and information attention duration?*

*Q2.3d*

*Is there a relationship between information acquisition styles and the degree of agreement on information access sequence and information attention duration?*

*Q2.3e*

*Is there a relationship between information processing styles and the degree of agreement on information access sequence and information attention duration?*

*Q2.3f*

*Is there a relationship between tolerance for ambiguity and the degree of agreement on information access sequence and information attention duration?*

#### **4.6.3 Auditors' materiality knowledge**

Sections 4.6.1 and 4.6.2 explained the development of a set of research questions aimed at describing auditors' problem-solving behavior at a high level of abstraction, in which general features of auditors' problem-solving behavior are related to psychological profiles and demographic characteristics. There remains one research question which is aimed at describing a specific auditor knowledge component - materiality assessment in the audit planning stage.

As has been said, expert system development is a means for learning more about how particular judgments are made [Steinbart, 1987]. In this dissertation knowledge about audit planning stage materiality is gathered for incorporation in an expert system. From studying that system, and from the process of its development, greater understanding is expected to be gained of the factors which determine planning stage materiality thresholds. The research question here is:

*Q3*

*What factors determine audit planning stage materiality thresholds?*

Auditor materiality knowledge is described at the knowledge level. This implies a description in three formats [Steels, 1992]: task structures, semantic networks and methods.

### Notes

1. Of the remaining 8 subjects, 2 had moved abroad and 1 had retired, leaving 5 subjects.
2. In order get feedback from all the subjects, we provided some substitute cases which were to be used if subjects did not bring in a case themselves. These substitute cases were based on the annual accounts of international businesses.
3. Since data is analyzed by means of nonparametric statistics, only the rank of this value is relevant.
4. This integral encoding took a total of 146 hours, an average of 5.84 hours per protocol.
5. Actually, after a subject's literal protocol was encoded into relevant statements by the research assistant, the encoding was integrally discussed with the experienced faculty member in order to reach consensus on the final encoding into relevant statements. Following this approach, a kappa coefficient was not calculated, nor was the percentage of agreement. However, an experience effect was noticed, in that consensus appeared to increase with the number of encoded protocols.
6. Because the number of row categories is not equal to the number of column categories, a Kappa coefficient could not be calculated. Instead, the raw agreement, which is slightly higher, is reported.
7. Subject number 5 has not been considered in this or any of the following analyses because he is an outlier, since his completeness ratio (that is the number of accessed pages divided by the total number of accessible pages) amounted 0.27, which is very low when compared to the group average (0.70). He only accessed the following information pages: assets and liabilities 1988, profit and loss accounts 1986 and 1988, results compliance tests and substantive tests, and audit values assets 1988. When he was asked to think aloud while solving the case, he immediately determined a planning stage materiality threshold on the basis of his firm's audit manual and the limited information which he had accessed. This process has been transcribed literally and coded into relevant statements and is summarized under the heading 'concurrent protocols'. Subsequently, he started a theoretical discussion about the differences between materiality and risk and the importance of quantitative materiality threshold assessment in auditing as well as in takeover price determination. This discussion has been transcribed literally and coded into relevant statements and is summarized under the heading 'retrospective'. So the data in table 7 for subject number 5 is not comparable with that for other subjects.

---

## Auditors' psychological profiles in relation to demographic characteristics

---

### 5.1 Introduction

This chapter and chapters 6 and 7 contain the empirical part of this dissertation. The purpose of this chapter is to provide evidence to justify further analysis of the demographic variables (number of years of audit experience, firm affiliation, formal education) and the personality variables (information acquisition style, information processing style, and tolerance for ambiguity). This analysis is done by investigating the distribution of the personality variables and testing for interrelationships between the personality variables and the demographic variables. The acquired data set will be discussed first, followed by a description of a set of research questions developed to augment the research questions in chapter 6 of this study. More specifically, these questions are posed in order to assess whether or not auditors' psychological profiles are related to their firm affiliation, number of years of auditing experience, and formal education. The general research question investigated here is:

*Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?*

In section 5.2, the results of the research questions which are investigated are discussed. More specifically, the scores on the demographic and personality variables are presented and analyzed in order to assess the potential for further analysis of these variables in relation to actual problem-solving behavior. In section 5.3, the findings are summarized and some conclusions are drawn.

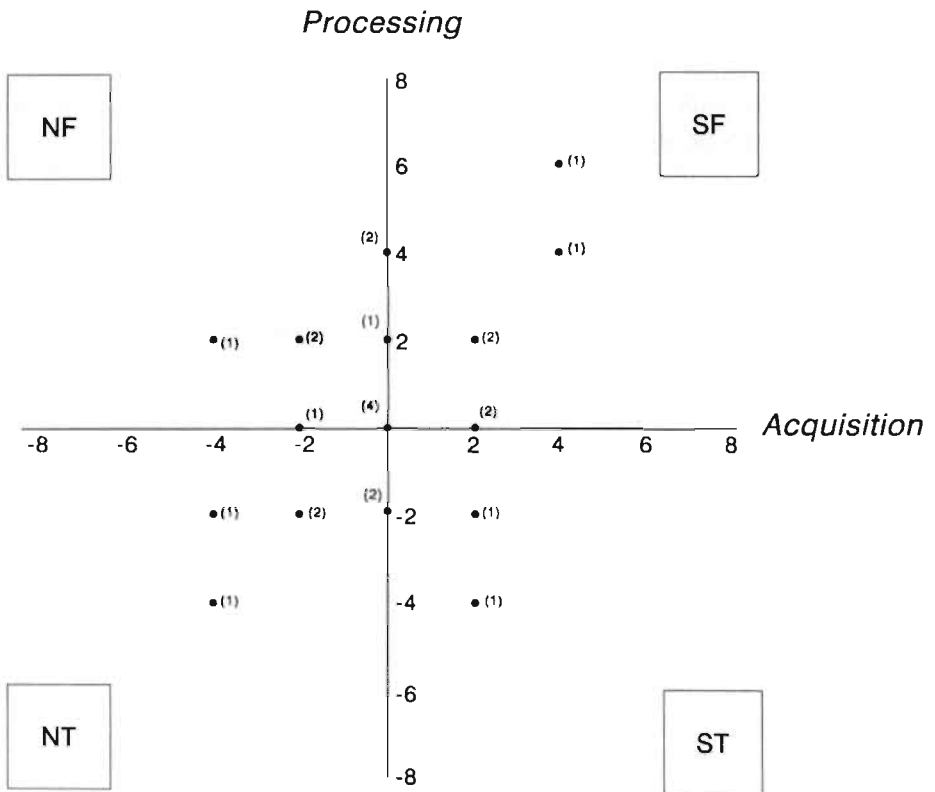
### 5.2 Results

#### 5.2.1 Psychological profiles

Research question 1.1a asks whether there are as many subjects with a sensing information acquisition style as with an intuitive acquisition style. Research question 1.1b asks whether there are as many subjects with a feeling information processing style as with a thinking processing style. Information acquisition and processing styles are measured by means of the Myers-Briggs type indicator.

The Myers-Briggs type indicator scores are graphed in figure 1. The first quadrant (upper right) contains the sensing/feeling (SF) types. The second quadrant (upper left) contains the intuitive/feeling (NF) types. The third quadrant (bottom left) contains the intuitive/thinking (NT) types. The fourth quadrant (bottom right) contains the sensing/thinking (ST) types. Some markers in figure 1 represent more than one observation. This is indicated by the numbers between brackets. For example, there are four subjects who scored zero on the information acquisition dimension and zero on the information processing dimension.

Figure 1: Scatter diagram of information acquisition and information processing scores.



Research question 1.1c investigates whether there are as many subjects with a high tolerance for ambiguity as with a low tolerance for ambiguity. Tolerance for ambiguity is measured by means of the MacDonald Tolerance for Ambiguity instrument [MacDonald, 1970].

In table 1, the numbers of subjects in each information acquisition, information processing, and tolerance for ambiguity category are cross-tabulated. It can be concluded that most auditors exhibit a high tolerance for ambiguity (research question 1.1c: 14 versus 6), but these auditors cannot as a group be classified into one of the information acquisition categories (research question 1.1a: 8 versus 8) or into one of the information processing categories (research question 1.1b: 8 versus 10).

Table 1: Numbers of subjects in each information acquisition-, information processing- and tolerance for ambiguity-category.

Information processing style	Tolerance for Ambiguity	Information acquisition style			
		Intuition	Neutral	Sensing	Total
Thinking	Tolerance low	1		1	2
	Tolerance neutral	1			1
	Tolerance high	2	2	1	5
	Total	4	2	2	8
Neutral	Tolerance low		3		3
	Tolerance neutral				
	Tolerance high	1	1	2	4
	Total	1	4	2	7
Feeling	Tolerance low		1		1
	Tolerance neutral	2		2	4
	Tolerance high	1	2	2	5
	Total	3	3	4	10
Total	Tolerance low	1	4	1	6
	Tolerance neutral	3		2	5
	Tolerance high	4	5	5	14
	Total	8	9	8	25

5.2.2 The relationships between demographics and psychological profiles

Research questions 1.2a, 1.2b and 1.2c investigate whether or not there is a relationship between experience and psychological profiles, as measured by the Myers-Briggs type indicator and the Tolerance for Ambiguity instrument. If experience is related to psychological profile this might indicate that psychological profiles are subject to change over time and hence are not stable subjective

characteristics but merely dispositions [Van de Poel, 1986, p. 104]. The subjects in this study had between 8 and 39 years of auditing experience, with an average of 21.40 (standard deviation: 8.85). Because most subjects were very experienced, and previous research [Ashton, 1974; Ashton & Kramer, 1980; Ashton & Brown, 1980; Hamilton & Wright, 1982; Biggs & Mock, 1980, 1983; Tabor, 1983; Frederick & Libby, 1986; Bedard, 1989] indicates that experience effects should be assessed within the range of 1 to 5 years of on-the-job experience, no significant experience effects were expected. However, Johnson, Jamal & Berryman [1989] found experience effects between a subject with 11 years of experience and an expert with 40 years of experience, indicating that certain tasks require more on-the-job experience than is normally realized.

Two approaches have been used in assessing the relationship between experience and psychological profiles. Firstly, the Pearson correlation coefficients<sup>1</sup> between the number of years of auditing experience and the tolerance for ambiguity and Myers-Briggs type indicator scores were calculated. Secondly, two experience categories were distinguished (variable name: *EXPCAT*): relatively inexperienced and relatively experienced. The cut-off point has been determined on the basis of the median number of years experience. Since, the median is 21 years, the variable *EXPCAT* is coded 1 for subjects who have more than 21 years of audit experience. For the remaining subjects, the variable *EXPCAT* is coded 0. An ANOVA procedure has been used to investigate whether or not there is a relationship between experience category and psychological profile.

It was found that the Pearson correlation coefficients between tolerance for ambiguity, information acquisition, information processing and the number of years auditing experience were:  $R_p = -0.075$ ,  $R_p = -0.036$ , and  $R_p = 0.109$  respectively. None of these correlation coefficients was significant. An ANOVA revealed F-values for tolerance for ambiguity, information acquisition, and information processing of: 0.041, 0.278, and 0.238, respectively. Here again, no significant effect was observed.

The question thus remains whether the relationship between psychological profile and experience is indeed non-existent, or whether experience effects cannot be observed if the minimum observed number of years of experience has exceeded a specific value. Of course, this question cannot be answered without gathering similar data for less experienced subjects and assessing the exact value of the 'critical number of years of experience'.

Research questions 1.3a, 1.3b and 1.3c investigate whether or not there is a relationship between educational background and psychological profiles, as measured by the Myers-Briggs type indicator and the tolerance for ambiguity instrument. There might be differences in psychological profiles between auditors



who have had an academic (university) or NIVRA education, in view of the divergent characteristics of the two educational categories (e.g. mainly full-time versus mainly part-time courses). In table 2, the average scores and significance levels for each personality variable are presented for both educational categories (based on independent t-tests).

**Table 2:** Comparison of educational categories with respect to tolerance for ambiguity and Myers-Briggs type indicator.

	Academic	NIVRA	p-value
Tolerance for ambiguity (scored for high tolerance)	2.222	3.000	0.764
Myers-Briggs type indicator for acquisition (scored for sensing)	-0.222	0.000	0.820
Myers-Briggs type indicator for processing (scored for feeling)	0.889	0.125	0.489

It was found that there is no significant relationship between educational background and psychological profile as measured by the Tolerance for Ambiguity and the Myers-Briggs type indicator scores.

Differences between accounting firms with regard to psychological profiles were assessed by means of ANOVA and independent t-tests. Because only two subjects were affiliated with firm 4, firm analyses in this section and in the next chapter focus on firms 1, 2, and 3, who provided 7, 8, and 8 subjects respectively. Since previous research has indicated that there are differences in the degree of structure between the audit firms in our sample [see e.g. Cushing & Loebbecke, 1986], and considering the characteristics of certain psychological profiles [e.g. Macintosh, 1985], it was expected that affiliates of the more structured firms would have a lower tolerance for ambiguity, a sensing information acquisition style and a thinking information processing style, and *vice versa*. For that reason a one-tailed test is applied.

From the ANOVA it was found that differences between firms are significant only with regard to tolerance for ambiguity ( $F=4.529$ ,  $p=0.024$ ). This difference in tolerance for ambiguity scores is caused by firm 1, where the subjects had a significantly higher average tolerance for ambiguity score than those in firms 2 and 3.

Table 3 summarizes the inter-firm comparison results (average scores and one-

tailed significance levels for differences between firms 1 and 2, firms 1 and 3, and firms 2 and 3 respectively). From the t-tests, It was found that there were differences between firm 1 on the one hand and firms 2 and 3 with regard to tolerance for ambiguity: firm 1 affiliates are significantly more tolerant of ambiguity than firm 2 and firm 3 affiliates. Furthermore, it was found that affiliates of firm 1 had a significantly more feeling information processing style than affiliates of firm 2. These findings are in accordance with expectations, since firm 1 can be labeled a relatively unstructured firm whereas firm 2 can be labeled a relatively structured firm [cf. Kinney, 1986].

**Table 3:** Inter-firm comparison with regard to tolerance for ambiguity and Myers-Briggs type indicator.

	Firm 1	Firm 2	Firm 3	p(1,2)	p(1,3)	p(2,3)
Tolerance for ambiguity (scored for high tolerance)	6.857	0.250	1.250	0.008	0.029	0.321
Myers-Briggs type indicator for acquisition (scored for sensing)	-0.571	0.000	0.250	0.321	0.258	0.421
Myers-Briggs type indicator for processing (scored for feeling)	1.429	0.000	0.250	0.083	0.202	0.432

### 5.3 Summary and conclusions

Table 4 presents a summary of the significant findings in this chapter.

The psychological profiles of the auditors in the sample are mainly characterized by a high tolerance for ambiguity (Q1.1c). However, within the sample ( $n=25$ ) six subjects had a low tolerance for ambiguity. This indicates that a search for interrelationships between tolerance for ambiguity on the one hand and problem-solving behavior, notwithstanding the observed preponderance of high tolerance for ambiguity, might reveal interesting results. Furthermore, there appears to be no significant relationship between the number of years of auditing experience and personality scores as measured by the MacDonald Tolerance for Ambiguity test (Q1.4).

Evidence as to subjects' information acquisition and processing styles is not conclusive (Q1.1a and Q1.1b). There also appears to be no significant relationship between the number of years of auditing experience and cognitive style scores as measured by the Myers-Briggs type indicator (Q1.2a and Q1.2b). This may be because all subjects are very experienced and additional years of auditing experience will not alter personality scores once a certain critical experience level

has been attained. Of course another conclusion would be that psychological profiles are not subject to change over time.

**Table 4:** Summary of significant findings regarding auditors' psychological profiles and the relationships between psychological profiles and demographics.

Research question	Test	Yes/No (where applicable $\alpha=0.10$ )	Findings		
			(a) Myers-Briggs type indicator: acquisition dimension	(b) Myers-Briggs type indicator: processing dimension	(c) Tolerance for ambiguity
<i>Q1.1 Do certain psychological profiles occur equally frequently among auditors?</i>	N/A	Yes	Sensing = 8 Intuition = 8	Feeling = 10 Thinking = 8	High = 14 Low = 6
<i>Q1.2 Is experience related to psychological profile?</i>	Correlation	No			
	ANOVA	No			
<i>Q1.3 Are there differences between auditors with a NIVRA and an academic education as to psychological profile?</i>	t-test	No			
<i>Q1.4 Are there differences between accounting firms as to psychological profiles of their affiliates?</i>	ANOVA	Yes	F-value = 0.231	F-value = 0.660	F-value = 4.529
	t-test	Yes	$p > 0.10$	$p < 0.10$ Firm 1 feeling score > firm 2 feeling score	$P < 0.10$ Firm 1 tolerance > firm 2 and firm 3 tolerance

Differences regarding subjects' psychological profiles between accounting firms were found only with regard to tolerance for ambiguity. This might indicate a (conscious or unconscious) personnel selection process based on a positive

valuation of tolerance for ambiguity within at least one of the firms in the sample (Q1.4c).

There is no significant difference between personality as measured by tolerance for ambiguity and Myers-Briggs type indicator scores for auditors with an academic education and for those with a NIVRA education (Q1.3a, Q1.3b and Q1.3c).

The analyses in this chapter are aimed mainly at finding evidence to justify further research using the variables mentioned in this chapter. If significant relationships were found, then this would have consequences for the analyses in chapter 6. An example will clarify this. Suppose all subjects were highly tolerant to ambiguity. In that case any analysis of differences in problem-solving behavior due to differences in tolerance for ambiguity would be meaningless because no differences would be found. The same holds if all the subjects had the same information acquisition or information processing styles. However, if there was evidence that firm affiliation and educational background were strongly related to certain psychological profiles, then this would indicate that no search for differences in problem-solving behavior due to differences in firm affiliation, educational background and psychological profiles should be undertaken in isolation, because of the observed covariability. Finally, if experience was strongly related to psychological profile, this would indicate that psychological profiles are not a stable personal characteristic. The findings in this analysis seem, on the whole, promising for further analyses aimed at assessing the relationships between personality and demographic variables on the one hand and actual problem-solving behavior on the other. Both the tolerance for ambiguity and the Myers-Briggs type indicator questionnaires can be used in the analyses.

### Notes

1. Based on the assumption that a 'yes' on one item within a questionnaire is equivalent to a 'yes' on any other item within the same questionnaire, an interval scale may be assumed. In the tolerance for ambiguity and Myers-Briggs type indicator questionnaires, each affirmative answer leads to a one-point score increase whereas a negative answer leads to a one-point score decrease.



## **Auditors' psychological profiles, demographic characteristics, and problem-solving behavior**

---

### **6.1 Introduction**

The purpose of this chapter is to determine whether there are relationships between auditors' psychological profiles and demographic characteristics on the one hand, and their problem-solving behavior in a planning stage materiality assessment case on the other hand. For that purpose, auditors' problem-solving behavior is measured by means of verbal protocol analysis (VPA) and automated process tracing (APT). The general research question investigated here is:

*Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?*

The operators and task-encodings which describe problem-solving behavior are derived from verbal protocols, which are analyzed to identify relationships between demographic factors (years of auditing experience, firm affiliation and educational background), psychological profiles (information acquisition style, information processing style and tolerance for ambiguity) and audit problem-solving behavior as measured by concurrent and retrospective protocol encodings and automated process traces.

In section 6.2, the results of the research questions which are investigated are discussed. In section 6.3, the findings are summarized and some conclusions are drawn.

### **6.2 Results**

#### **6.2.1 Operators and task encodings**

Research question 2.1a investigates whether or not there is a relationship between the relative number of operators in each of the categories of information search, information retrieval, judgment and decision, and the number of years auditing experience. The direction of the relationship, if any, cannot be predicted, so a two-tailed test is applied. The variables *IS*, *IR*, *J* and *D* represent absolute numbers of operators assigned to relevant statements. However the relative number of operators is a more suitable measure, because the total number of relevant statements differs between subjects. The relative numbers of operators per subject

are computed as:  $IS/NRS$ ,  $IR/NRS$ ,  $J/NRS$  and  $D/NRS$ , where  $NRS$  is the number of relevant statements per subject. Because the resulting variables are ratio-scaled, independent t-tests are used. Table 1 summarizes the results.

**Table 1:** Results of independent t-tests on the relationship between relative operator frequencies and experience.

Experience category	Relative operator frequencies			
	Information search ( $IS/NRS$ )	Information retrieval ( $IR/NRS$ )	Judgment ( $J/NRS$ )	Decision ( $D/NRS$ )
Less experienced subjects ( $EXPCAT=0$ )	0.1141	0.4830	0.3847	0.0182
More experienced subjects ( $EXPCAT=1$ )	0.0613	0.5783	0.3485	0.0119
p-value	0.087	0.088	0.545	0.450

It was found that the more experienced subjects apply significantly less information search operators and significantly more information retrieval operators, when compared to the less experienced subjects. This indicates that experienced subjects needed less information as input to their decision processes. However, the experienced subjects, as a consequence of 'on-the-job' learning, have stored more information in memory and hence can retrieve more information, which they apparently did.

Research question 2.1b investigates whether or not there is a relationship between the relative number of operators in each of the categories of information search, information retrieval, judgment and decision, and educational background. Here again, the direction of the relationship cannot be predicted, so a two-tailed test is applied. The variable for education category,  $EDU$ , is coded 1 for subjects who have an academic education and 0 for subjects who have a NIVRA education. In table 2 the results are summarized.

It was found that the subjects with a NIVRA education apply significantly more information search operators when compared to subjects with an academic education. This indicates that NIVRA-educated auditors needed more information as input to their decision processes. No significant differences with respect to the categories of information retrieval, judgment and decision were observed.

Research question 2.1c investigates whether or not there is a relationship between



the relative number of operators in each of the categories of information search, information retrieval, judgment and decision, and audit firm affiliation. Here again, the direction of the relationship could not be predicted, so a two-tailed test is applied. Four audit firms participated in the research project. However, because firm 4 provided only two subjects the current analyses focus on firms 1 to 3 (variable name: *FIRM*). In order to investigate differences between firms, an ANOVA was first applied. Table 3 summarizes the results.

**Table 2:** Results of independent t-tests on the relationship between relative operator frequencies and educational background.

Education category	Relative operator frequencies			
	Information search ( <i>IS/NRS</i> )	Information retrieval ( <i>IR/NRS</i> )	Judgment ( <i>J/NRS</i> )	Decision ( <i>D/NRS</i> )
NIVRA education ( <i>EDU=0</i> )	0.1132	0.5140	0.3565	0.0164
Academic education ( <i>EDU=1</i> )	0.0510	0.5479	0.3875	0.0136
p-value	0.025	0.566	0.615	0.751

**Table 3:** Results of ANOVA on the relationship between relative operator frequencies and audit firm affiliation.

Firm affiliation	Relative operator frequencies			
	Information search ( <i>IS/NRS</i> )	Information retrieval ( <i>IR/NRS</i> )	Judgment ( <i>J/NRS</i> )	Decision ( <i>D/NRS</i> )
Between groups sum of squares	0.0068	0.0727	0.0655	0.0024
Within groups sum of squares	0.1208	0.2682	0.2895	0.0062
p-value	0.592	0.102	0.144	0.042

Since the p-values for *IR/NRS* and *J/NRS* are greater than 0.10 but nevertheless low, and for *D/NRS* the p-value is less than 0.10, further analyses as to individual firm differences were made on the basis of independent t-tests. The results are summarized in table 4.

**Table 4:** Results of independent t-tests on the relationship between relative operator frequencies and audit firm affiliation.

Firm affiliation	Relative operator frequencies			
	Information search ( <i>IS/NRS</i> )	Information retrieval ( <i>IR/NRS</i> )	Judgment ( <i>J/NRS</i> )	Decision ( <i>D/NRS</i> )
FIRM=1	0.0807	0.6282	0.2816	0.0094
FIRM=2	0.1180	0.5111	0.3403	0.0306
FIRM=3	0.0796	0.5000	0.4134	0.0069
p-value differences between firm 1 and 2	0.459	0.109	0.387	0.076
p-value differences between firm 1 and 3	0.974	0.053	0.048	0.708
p-value differences between firm 2 and 3	0.412	0.855	0.300	0.041

It was found that firm 1 and firm 3 subjects have a significantly lower average *D/NRS* than firm 2 subjects. This indicates that firm 2 subjects are more decisive with respect to materiality decisions than firm 1 and firm 2 subjects. Furthermore it was found that firm 1 subjects have a significantly higher average *IR/NRS* and a significantly lower average *J/NRS* than firm 3 subjects. No significant differences were observed, between firms, with respect to the average *IS/NRS*.

Research question 2.1d investigates whether or not there is a relationship between the relative number of operators in each of the categories of information search, information retrieval, judgment and decision, and information acquisition style. The sensing types would be expected to have a preponderance of information search operators, so a one-tailed test is applied. With respect to the information retrieval, judgment and decision operators no specific direction is expected, so a two-tailed test is applied. In table 5 the results are summarized.

It was found that the subjects with a sensing information acquisition style apply significantly more information search operators when compared to the subjects with an intuitive information acquisition style. In accordance with expectations, no significant differences were observed with respect to the categories information retrieval, judgment, and decision.

**Table 5:** Results of independent t-tests on the relationship between relative operator frequencies and information acquisition style.

Information acquisition style	Relative operator frequencies			
	Information search (IS/NRS) (1-tail sign.)	Information retrieval (IR/NRS)	Judgment (J/NRS)	Decision (D/NRS)
Intuitive style	0.0717	0.5155	0.4002	0.0126
Sensing style	0.1262	0.5300	0.3241	0.0197
p-value	0.086	0.855	0.312	0.505

Research question 2.1e investigates whether or not there is a relationship between the relative number of operators in each of the categories of information search, information retrieval, judgment and decision, and information processing style. Thinking types would be expected to have a preponderance of information retrieval operators, so a one-tailed test is applied. No specific direction is expected with respect to the information search, judgment and decision operators, so a two-tailed test is applied. Furthermore, feeling types would be expected to have a significantly lower relevance ratio (word count in relevant statements divided by word count in the literal transcripts: *WCRS/WCPT*), so a one-tailed test is applied. Finally, thinking types were expected to take significantly more time (*TT*) for decision-making. So, here again a one-tailed test is applied. Table 6 summarizes the results.

It was found that the subjects with a feeling information processing style apply significantly more information retrieval operators than subjects with a thinking information processing style ( $p\text{-value}=0.023$ ). This is an unexpected finding, since information retrieval processes are expected to occur more frequently among thinking types than feeling types. As expected, no significant differences were observed with respect to the categories information search, judgment and decision. Also in accordance with expectations, feeling types have a significantly lower relevance ratio than thinking types ( $p\text{-value}=0.050$ ). Finally, and also in accordance with expectations, thinking types need significantly more time to solve the case ( $p\text{-value}=0.065$ ).

Research question 2.1f1 investigates whether or not there is a relationship between tolerance for ambiguity and the relative number of operators in each of the categories of information search, information retrieval, judgment and decision. Subjects with a low tolerance for ambiguity would be expected to have a preponderance of judgment and decision operators, and subjects with a high tolerance for ambiguity were expected to have a preponderance of information

search and information retrieval operators. For that reason a one-tailed test was applied. In table 7 the results are summarized.

**Table 6:** Results of independent t-tests on the relationship between relative operator frequencies, total time and relevance ratio, and information processing style.

Information processing style	Relevance ratio ( <i>RR</i> ) (1-tail sign.)	Total time ( <i>TT</i> ) (1-tail sign.)	Relative operator frequencies			
			Information search ( <i>IS/NRS</i> )	Information retrieval ( <i>IR/NRS</i> ) (1-tail sign.)	Judgment ( <i>J/NRS</i> )	Decision ( <i>D/NRS</i> )
Thinking style	0.2316	3468	0.1076	0.4305	0.4410	0.0210
Feeling style	0.1831	2734	0.0564	0.5514	0.3758	0.0164
p-value	0.050	0.065	0.117	0.023	0.328	0.678

No significant results were found. This could be because most subjects had a high tolerance for ambiguity, so differences with respect to any variable due to differences in degree of tolerance for ambiguity would be obscured. However, evidence as to the degree of tolerance for ambiguity is not conclusive (see chapter 5). Another way of examining the relationship between tolerance for ambiguity and problem-solving behavior is to ask whether subjects show which task is being addressed (research question 2.1f2). Here, subjects with a high tolerance for ambiguity would be expected to verbalize more audit operators, more takeover price operators, more explicitly determined operators, and less undetermined operators. For that reason a one-tailed test is applied. The results are summarized in table 8.

No significant results were found. Again, this could be because most subjects had a low tolerance for ambiguity. However the direction of the outcomes is as expected: subjects with a low tolerance for ambiguity had a higher average number of determined operators (0.4410 versus 0.3756), whereas subjects with a high tolerance for ambiguity had a higher average number of undetermined operators (0.6244 versus 0.5590).

**Table 7:** Results of independent t-tests on the relationship between relative operator frequencies and tolerance for ambiguity

Tolerance for ambiguity	Relative operator frequencies			
	Information search (IS/NRS) (1-tail sign.)	Information retrieval (IR/NRS) (1-tail sign.)	Judgment (J/NRS) (1-tail sign.)	Decision (D/NRS) (1-tail sign.)
Low tolerance for ambiguity	0.0883	0.4931	0.4093	0.0093
High tolerance for ambiguity	0.1066	0.5426	0.3372	0.0137
p-value	0.334	0.255	0.165	0.300

**Table 8:** Results of independent t-tests on the relationship between relative task frequencies and tolerance for ambiguity.

Tolerance for ambiguity	Relative task frequencies				
	Audit (A/NRS) (1-tail sign.)	Takeover price (T/NRS) (1-tail sign.)	Determined (DT/NRS) (1-tail sign.)	(A + T + DT)/NRS (1-tail sign.)	Indeterminate I/NRS (1-tail sign.)
Low tolerance for ambiguity	0.1948	0.1751	0.0711	0.4410	0.5590
High tolerance for ambiguity	0.1386	0.1425	0.0946	0.3756	0.6244
p-value	0.123	0.224	0.241	0.129	0.129

### 6.2.2 Recall and completeness of information access

Research question 2.2a investigates whether or not there is a relationship between years of auditing experience and the number of recalled relevant statements after correction for information access. A correction for information access is made on the basis of the variable: 'completeness of information access' (*COMP*) which measures the relative number of information pages that are accessed at least once during the concurrent protocol sessions. Two information access categories are distinguished (variable name: *COMPCAT*): low information access and high information access. The cut-off point has been determined on the basis of the median of *COMP*. Since the median is 0.69, the variable *COMPCAT* is coded 1 for subjects who accessed more than 69% of the information pages at least once, and 0 for the remaining subjects. In order to investigate the main effects and the

interaction effects of experience and completeness of information access on the number of recalled relevant statements, an ANOVA was performed. An interaction effect between experience and completeness of information access is expected, because experienced subjects who access all the information pages are expected to process that information effectively so that it can easily be recalled (e.g. Weber [1980], Boshuizen [1989], Christ [1993]). On the other hand, inexperienced subjects who access all the information pages are not expected to process that information effectively and consequently would not easily be able to recall that information [see e.g. Glazer *et al.*, 1992]. In table 9 the results are summarized.

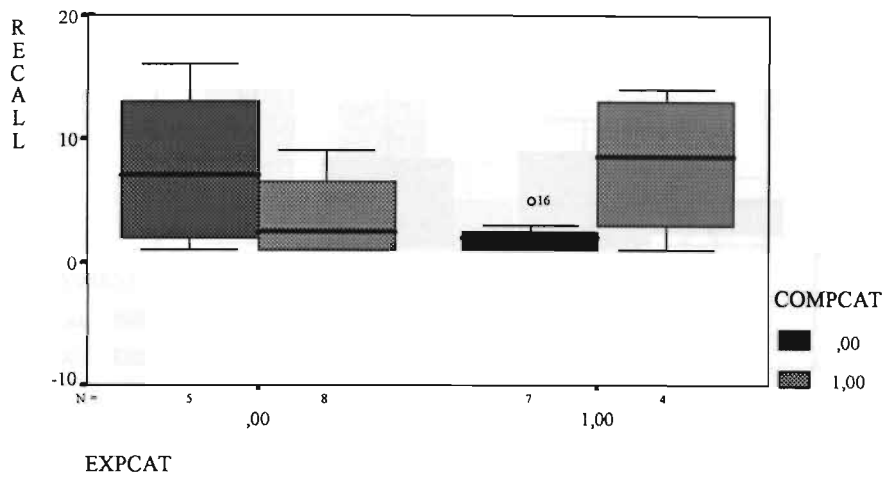
It appears that firm 1 affiliates and firm 3 affiliates who accessed more information recalled more than their counterparts who accessed less information. However, this finding is not significant: an independent t-test revealed that for firm 1 and firm 3 affiliates the average number of recalled relevant statements was 4.000 (*COMPCAT*=0) versus 6.800 (*COMPCAT*=1) with a p-value of 0.331. Firm 2 affiliates who accessed more information recalled less than their counterparts who accessed less information. However, this finding is again not significant: an independent t-test revealed that for firm 2 affiliates the average number of recalled relevant statements was 7.000 (*COMPCAT*=0) versus 3.600 (*COMPCAT*=1) with a p-value of 0.671.

**Table 9:** Results of ANOVA on the relationship between the number of recalled relevant statements, completeness of information access and experience.

Experience and information access categories	F-value	p-value
Main effects of <i>EXPCAT</i>	0.149	0.704
Main effects of <i>COMPCAT</i>	0.245	0.626
Interaction effects of <i>EXPCAT</i> and <i>COMPCAT</i>	7.368	0.013

A significant interaction effect was found between experience and information access ( $p=0.013$ ). To determine whether or not the direction of the interaction effect of experience and completeness of information access is as expected, the Pearson correlation coefficients between the number of recalled relevant statements (*RECALL*) and the completeness of information access (*COMP*) are calculated for both experience categories. It was found that the relatively inexperienced group had a significant negative correlation between *COMP* and *RECALL* ( $R_p=-0.5757$ ,  $p=0.020$ ), whereas the relatively experienced group had a significant positive correlation between *COMP* and *RECALL* ( $R_p=0.8199$ ,  $p=0.001$ ). In figure 1, the relationship between *RECALL*, *EXPCAT* and *COMPCAT* is visualized by means of a boxplot.<sup>1</sup>

Figure 1: Boxplot of recall, experience and completeness of information access.



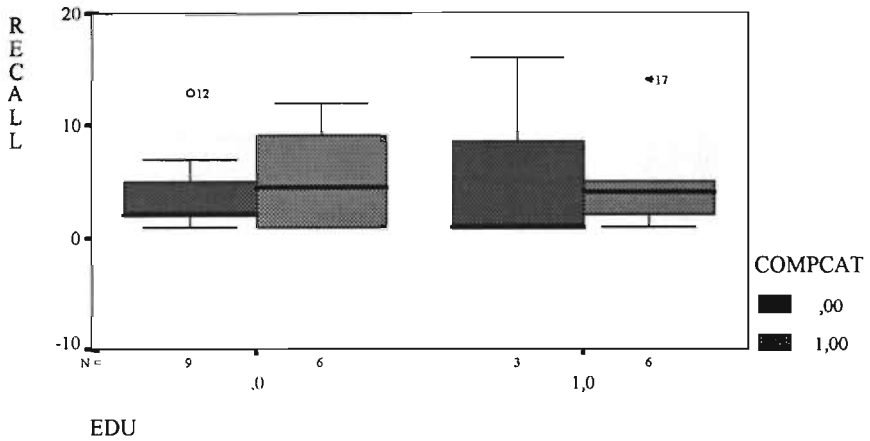
It appears that, among the inexperienced subjects, those who accessed more information recalled less than those who accessed less information. On the other hand, for experienced subjects, those who accessed more information recalled more than those who accessed less information.

Research question 2.2b investigates whether or not there is a relationship between the educational background and the number of recalled relevant statements after correction for information access. In order to investigate the main effects and interaction effects of educational background (*EDU*) and completeness of information access on the number of recalled relevant statements, an ANOVA was performed. No interaction effect is expected between educational background and completeness of information access. In table 10 the results are summarized.

Table 10: Results of ANOVA on the relationship between the number of recalled relevant statements, completeness of information access and educational background.

Education and information access categories	F-value	p-value
Main effects of <i>EDU</i>	0.142	0.710
Main effects of <i>COMPCAT</i>	0.006	0.941
Interaction effects of <i>EDU</i> and <i>COMPCAT</i>	0.278	0.604

Figure 2: Boxplot of recall, education and completeness of information access.



No main effects or interaction effects, between educational background and completeness of information access, were found. Figure 2 shows the relationship between *RECALL*, *EDU* and *COMPCAT* by means of a boxplot.

It appears that NIVRA-educated (*EDU*=0) and university-educated auditors (*EDU*=1) who accessed more information recalled more than their counterparts who accessed less information.

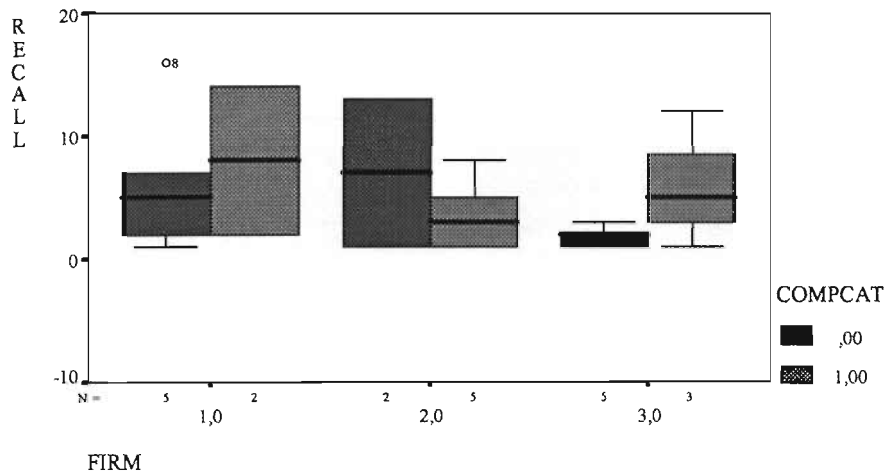
Research question 2.2c investigates whether or not there is a relationship between firm affiliation and the number of recalled relevant statements after correction for information access. In order to investigate the main effects as well as the interaction effects of firm affiliation (*FIRM*) and completeness of information access on the number of recalled relevant statements, an ANOVA was performed. No interaction effect would be expected between firm affiliation and completeness of information access. The results are summarized in table 11.

Table 11: Results of ANOVA on the relationship between the number of recalled relevant statements, completeness of information access and firm affiliation.

Firm and information access categories	F-value	p-value
Main effects of <i>FIRM</i>	0.687	0.517
Main effects of <i>COMPCAT</i>	0.145	0.709
Interaction effects of <i>FIRM</i> and <i>COMPCAT</i>	0.985	0.395



Figure 3: Boxplot of recall, firm affiliation and completeness of information access.



No main effects or interaction effects were found between firm affiliation and completeness of information access. Figure 3 shows the relationship between *RECALL*, *FIRM* and *COMPCAT* by means of a boxplot.

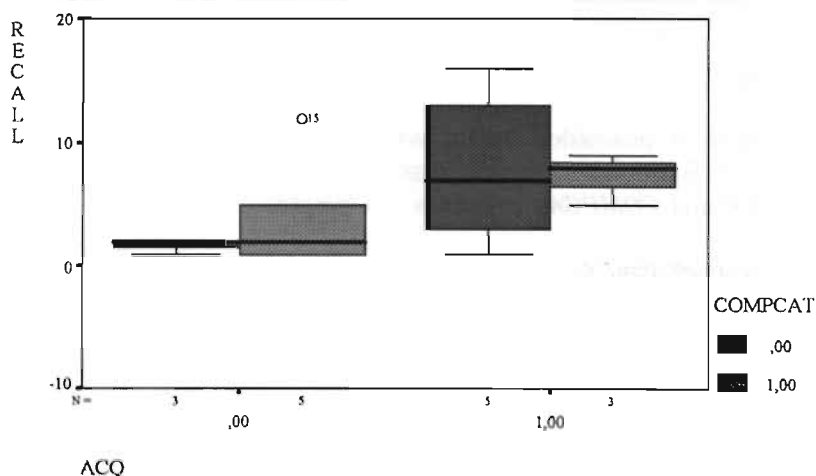
Firm 1 affiliates and firm 3 affiliates who accessed more information recalled more than their counterparts who accessed less information. However, this finding is not significant: an independent t-test revealed that for firm 1 and firm 3 affiliates the average number of recalled relevant statements was 4.000 (*COMPCAT*=0) versus 6.800 (*COMPCAT*=1) with a p-value of 0.331. Firm 2 affiliates who accessed more information recalled less than their counterparts who accessed less information. However, this finding is again not significant: an independent t-test revealed that for firm 2 affiliates the average number of recalled relevant statements was 7.000 (*COMPCAT*=0) versus 3.600 (*COMPCAT*=1) with a p-value of 0.671.

Research question 2.2d investigates whether or not there is a relationship between information acquisition style and the number of recalled relevant statements, after correction for information access. In order to investigate the main effects and interaction effects of information acquisition style (*ACQ*) and completeness of information access on the number of recalled relevant statements, an ANOVA was performed. No interaction effect would be expected between information acquisition style and completeness of information access. The results are summarized in table 12.

**Table 12:** Results of ANOVA on the relationship between the number of recalled relevant statements, completeness of information access and information acquisition style.

Information acquisition style and information access categories	F-value	p-value
Main effects of <i>ACQ</i>	3.876	0.073
Main effects of <i>COMPCAT</i>	0.151	0.705
Interaction effects of <i>ACQ</i> and <i>COMPCAT</i>	0.443	0.518

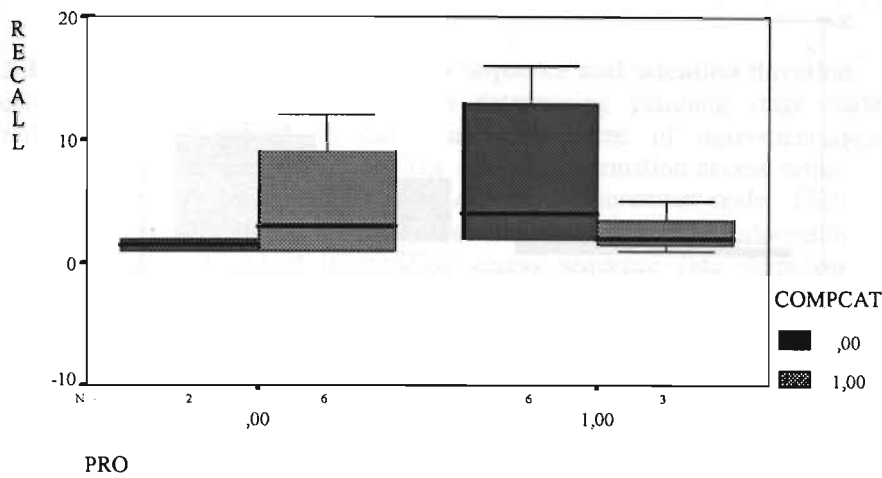
**Figure 4:** Boxplot of recall, information acquisition style and completeness of information access.



There is a main effect of information acquisition style ( $p=0.073$ ), but there was no main effect of completeness of information access and no interaction effect between information acquisition style and completeness of information access. Figure 4 shows the relationship between *REC ALL*, *ACQ* and *COMPCAT* by means of a boxplot. Sensing types ( $ACQ=1$ ), independent of the completeness of their information access, recalled more than intuitive types ( $ACQ=0$ ).

Research question 2.2e asks whether or not there is a relationship between information processing style and the number of recalled relevant statements after correction for information access. To investigate the main effects as well as the interaction effects of information processing style (*PRO*) and completeness of information access on the number of recalled relevant statements, an ANOVA was performed. No interaction effect was expected between information processing style and completeness of information access. The results are summarized in table 13.

Figure 5: Boxplot of recall, information processing style and completeness of information access.



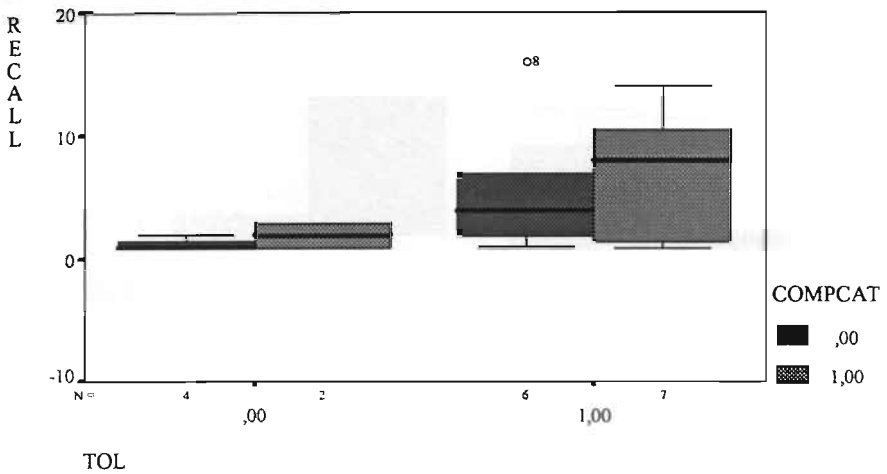
No main effect of information processing style ( $p=0.558$ ) was found. Nor is there an interaction effect between information processing style and completeness of information access. There was also no main effect of completeness of information access, and no interaction effect between information acquisition style and completeness of information access. Figure 5 shows the relationship between *RECALL*, *PRO* and *COMPCAT* by means of a boxplot.

Table 13: Results of ANOVA on the relationship between the number of recalled relevant statements, completeness of information access and information processing style.

Information processing style and information access categories	F-value	p-value
Main effects of <i>PRO</i>	0.362	0.558
Main effects of <i>COMPCAT</i>	0.025	0.877
Interaction effects of <i>PRO</i> and <i>COMPCAT</i>	2.030	0.178

Feeling types ( $PRO=1$ ) who accessed more information recalled less than feeling types who accessed less information and thinking types ( $PRO=0$ ) who accessed more information recalled more than thinking types who accessed less information. However, this difference is not significant ( $p=0.178$ ).

Research question 2.2f investigates whether there is a relationship between

**Figure 6:** Boxplot of recall, tolerance for ambiguity and completeness of information access.

tolerance for ambiguity and the number of recalled relevant statements, after correction for information access. In order to investigate the main effects and interaction effects of information processing style (*TOL*) and completeness of information access on the number of recalled relevant statements, an ANOVA was performed. No interaction effect was expected between tolerance for ambiguity and completeness of information access. Table 14 summarizes the results.

**Table 14:** Results of ANOVA on the relationship between the number of recalled relevant statements, completeness of information access and tolerance for ambiguity.

Tolerance for ambiguity and information access categories	F-value	p-value
Main effects of <i>TOL</i>	3.582	0.078
Main effects of <i>COMPCAT</i>	0.139	0.715
Interaction effects of <i>TOL</i> and <i>COMPCAT</i>	0.004	0.952

There is a main effect of tolerance for ambiguity ( $p=0.078$ ), but no interaction effect between tolerance for ambiguity and completeness of information access. Figure 6 shows the relationship between *RECALL*, *TOL* and *COMPCAT* by means of a boxplot.

Subjects with a high tolerance for ambiguity ( $TOL=1$ ) recalled significantly more ( $p=0.078$ ) than subjects with a low tolerance for ambiguity ( $TOL=0$ ), independent

of the completeness of information access.

6.2.3 Agreement on information access sequence and attention duration

Auditors' problem-solving behavior in determining planning stage materiality thresholds was examined to determine the degree of agreement regarding information page access sequences. The data on information access sequences was transformed into ranks, resulting in an ordinal measurement scale. Therefore the Spearman correlation coefficient ( $R_s$ ) between each subject's information access sequence and the standard information access sequence (see Appendix I) was calculated. It was found that 4 out of 24 subjects followed a sequential information access strategy (that is, the menu-induced sequence, top-down or bottom-up). To determine overall agreement regarding compliance with the standard information access sequence, the average  $R_s$  ( $\text{ave}[R_s]$ ) was calculated based on Fisher's-Z transformation<sup>2</sup> (see table 15). It was found that  $\text{ave}[R_s]=0.0141$ , which is not significant for  $\alpha=0.10$ . This indicates that there is no agreement with respect to the standard information access sequence.

Table 15: Spearman correlation coefficients for observed information access sequence and standard information access sequence.

Subject	$R_s$	*: $p < 0.10$ **: $p < 0.01$	Subject	$R_s$	*: $p < 0.10$ **: $p < 0.01$
1	0.0268		14	-0.0717	
2	-0.3161		15	0.4328	*
3	-0.0346		16	-0.3259	
4	-0.0215		17	0.5813	**
5	N/A		18	-0.3912	*
6	0.3274		19	-0.1170	
7	-0.2265		20	-0.2213	
8	-0.0839		21	-0.2222	
9	-0.1595		22	0.1284	
10	0.4978	**	23	0.1046	
11	0.1045		24	0.0756	
12	0.0004		25	-0.1689	
13	0.3208		Ave[ $R_s$ ]	0.0141	

In spite of the above results with regard to auditors following our standard access sequence, there could be agreement among auditors around some other access sequence. The standard access sequence used in this research is highly subjective and might not be the sequence generally chosen by auditors. In order to assess agreement more objectively, the Spearman  $R$  between each subject's information access sequence and the average information access sequence of the other 23 subjects is first calculated as a measure of agreement (see table 16). Here again, the average Spearman correlation coefficient is calculated based on Fisher's  $Z$  transformation. It was found that  $\text{ave}[R_s] = 0.5429$ , which is significant for  $\alpha = 0.01$ . Considering the number of significant correlation coefficients as represented in table 16, there is an underlying ranking on which most subjects (18 out of 24) agree.

**Table 16:** Spearman correlation coefficients for observed information access sequence and consensus information access sequence.

Subject	$R_s$	*: $p < 0.10$ **: $p < 0.01$	Subject	$R_s$	*: $p < 0.10$ **: $p < 0.01$
1	0.5635	**	14	0.5113	**
2	0.7207	**	15	0.3090	
3	0.5117	**	16	0.7133	**
4	0.6539	**	17	-0.0216	
5	N/A		18	0.6621	**
6	0.3465	*	19	0.8116	**
7	0.5207	**	20	0.3802	*
8	0.5425	**	21	0.8187	**
9	0.7146	**	22	0.5417	**
10	-0.0162		23	0.3003	
11	0.2767		24	0.2990	
12	0.8044	**	25	0.7287	**
13	0.4586	*	Ave[ $R_s$ ]	0.5429	**

In that strategy, the most recent quantitative pages concerning the balance sheet and profit and loss accounts are accessed first, followed by the valuation, investments, management and contracts pages. Furthermore, the profit and loss account for 1986 is accessed. Presumably, auditors want to construct a picture about the trend in income components.<sup>3</sup> As might be expected, the page dealing with the setup of

internal controls is accessed before the page giving the results of compliance tests, which is accessed before the results of the substantive tests. The audit value pages are relatively highly ranked (15, 20 and 22, see Appendix I), which indicates their position within the audit process. In general, from the average ranking observed, it can be concluded that auditors mainly follow the textbook (or audit manual) access sequence [Jenkins *et al.*, 1986; Arens & Loebbecke, 1991].

A quick glance through tables 15 and 16 would lead one to expect that the Spearman correlation coefficient between the standard information access sequence and the average information access sequence would be negative. It was found that this is indeed the case:  $R_s = -0.2014$ . However, the significance is only  $\alpha = 0.162$ . Nevertheless, it might be that in the following analyses significant relationships between measures of agreement with the information access sequences (SS and CS) and demographic or personality variables are in an unexpected direction, since expected results for agreement on standard sequence would probably lead to unexpected results for agreement on average sequence and *vice versa*. This should be kept in mind when analyzing the differences between the different categories defined by experience, educational background, firm affiliation, information acquisition and information processing styles, and tolerance for ambiguity categories.

The degree of agreement among the subjects as to the time to spend on each information page was assessed in order to find evidence which might be useful in setting time budgets for audit planning. The same method as applied in examining information access strategies was used. Firstly, the Spearman R between each subject's information attention duration per information page and the average information attention duration of the other 23 subjects was calculated to assess agreement on an individual level. It appeared that correlation was significant for 22 out of 24 subjects. Secondly, the average  $R_s$  was calculated based on Fisher's Z transformation in order to determine overall agreement regarding information page attention duration (see table 17). It was found that  $\text{ave}[R_s] = 0.6746$ , which is significant for  $\alpha = 0.01$ .

Appendix I also presents the average attention duration per information page. In the average pattern of time expenditure, the information pages which received the most attention were the profit and loss account and the balance sheet of the year 1988. Much time was also spent on market expectations, the results of compliance tests, the audit value pages, and on contract information, perhaps because of the ambiguous nature of the information represented on these pages. Little attention was paid to financial information for the years 1985, 1986 and 1987.

**Table 17:** Spearman correlation coefficients for observed information attention duration and average information attention duration.

Subject	$R_s$	*: $p < 0.10$ **: $p < 0.01$	Subject	$R_s$	*: $p < 0.10$ **: $p < 0.01$
1	0.6715	**	14	0.6933	**
2	0.6178	**	15	0.6682	**
3	0.6558	**	16	0.8048	**
4	0.6965	**	17	0.5792	**
5	N/A		18	0.7845	**
6	0.5375	**	19	0.7719	**
7	0.9084	**	20	0.6554	**
8	0.8822	**	21	0.5497	**
9	0.6321	**	22	0.6652	**
10	0.1870		23	0.0520	
11	0.5593	**	24	0.6170	**
12	0.6807	**	25	0.6815	**
13	0.8484	**	Ave[ $R_s$ ]	0.6746	**

Research question 2.3a asks whether there is a relationship between the number of years auditing experience and the degree of agreement with the information access sequences (*SS* and *CS*) and average information attention duration (*AD*). Independent t-tests were used to investigate any such differences. Table 18 shows the results: no specific effects were expected.

It was found that the more experienced subjects show significantly ( $p\text{-value}=0.031$ ) less agreement with the average attention duration per information page than the less experienced subjects. The more experienced subjects also show less agreement with the average information access sequence than the less experienced subjects. However, this difference is not significant ( $p\text{-value}=0.140$ ).

Research question 2.3b investigates whether there is a relationship between educational background and the degree of agreement with the information access sequences (*SS* and *CS*) and average information attention duration (*AD*). Independent t-tests were performed to investigate any such differences. Here again, no specific effects are expected, and no significant effects were found. However, auditors with an academic education ( $EDU=1$ ) showed more agreement ( $p\text{-value}=0.201$ ) with the average information attention duration than auditors with a



NIVRA education (*EDU*=0). Table 19 summarizes the results.

**Table 18:** Results of independent t-tests on the relationship between degree of agreement with information access sequences and average information attention duration and experience.

Experience category	Agreement		
	Standard sequence ( <i>SS</i> )	Average sequence ( <i>CS</i> )	Attention duration ( <i>AD</i> )
Less experienced subjects ( <i>EXPCAT</i> =0)	-0.0227	0.5906	0.7173
More experienced subjects ( <i>EXPCAT</i> =1)	0.0445	0.4251	0.5523
p-value	0.568	0.140	0.031

Research question 2.3c asks whether there is a relationship between firm affiliation and the degree of agreement with the information access sequences and average information attention duration. Once again, independent t-tests were performed, with no specific effects being expected. The results are summarized in table 20.

No significant differences were found. An ANOVA revealed F-ratios for *SS*, *CS* and *AD* of 0.1642, 1.0901 and 0.6041, respectively, none of which are significant for  $\alpha=0.10$ . Thus there was no relationship between firm affiliation and the degree of agreement with the information access sequences or average information attention duration.

Research question 2.3d asks whether there is a relationship between information acquisition style and the degree of agreement with the information access sequences and average information attention duration. Sensing types would be expected to show more agreement with the average sequence (or the standard sequence) and average attention duration than intuitive types, because they are more fact-oriented, less creative, and emphasize precision rather than global values [Keen & Bronsema, 1981]. Therefore a one-tailed test is applied. Table 21 summarizes the results.

It was found that the subjects with a sensing information acquisition style show significantly more agreement with the average information access sequence than the subjects with an intuitive information acquisition style (p-value=0.010). No significant differences for information attention duration were found, but the direction of the differences was as expected.

**Table 19:** Results of independent t-tests on the relationship between degree of agreement with the information access sequences and average information attention duration and educational background.

Educational background	Agreement		Attention duration (AD)
	Standard sequence (SS)	Average sequence (CS)	
NIVRA education (EDU=0)	0.0403	0.5262	0.6027
Academic education (EDU=1)	-0.0456	0.4956	0.7067
p-value	0.452	0.782	0.201

**Table 20:** Results of independent t-tests on the relationship between degree of agreement with the information access sequences and average information attention duration and audit firm affiliation.

Firm affiliation	Agreement		Attention duration (AD)
	Standard sequence (SS)	Average sequence (CS)	
FIRM=1	0.0742	0.3981	0.6268
FIRM=2	-0.0037	0.5968	0.6995
FIRM=3	0.0049	0.5214	0.5861
p-value differences between firm 1 and 2	0.635	0.209	0.475
p-value differences between firm 1 and 3	0.655	0.414	0.733
p-value differences between firm 2 and 3	0.948	0.477	0.266

Research question 2.3e asks whether there is a relationship between information processing style and the degree of agreement with the information access sequences and average information attention duration. Thinking types would be expected to show more agreement with the average sequence (or the standard sequence) and average attention duration than feeling types, because they apply more rational information processing [Keen & Bronsema, 1981], so a one-tailed test is applied. It was found that subjects with a thinking information processing style were significantly more in agreement with the standard information access sequence than subjects with a feeling information processing style ( $p\text{-value}=0.098$ ). No significant differences for information attention duration were found, and the

direction of the differences found was contrary to what was expected. The results are summarized in table 22.

**Table 21:** Results of independent t-tests on the relationship between degree of agreement with the information access sequences and average information attention duration and information acquisition style.

Information acquisition style	Agreement		
	Standard sequence (SS) (1-tail sign.)	Average sequence (CS) (1-tail sign.)	Attention duration (AD) (1-tail sign.)
Intuitive style	0.0537	0.3842	0.6151
Sensing style	-0.0438	0.6578	0.6820
p-value	0.207	0.010	0.213

**Table 22:** Results of independent t-tests on the relationship between degree of agreement with the information access sequences and average information attention duration and information processing style.

Information processing style	Agreement		
	Standard sequence (SS) (1-tail sign.)	Average sequence (CS) (1-tail sign.)	Attention duration (AD) (1-tail sign.)
Thinking style	0.0472	0.5166	0.6493
Feeling style	-0.0807	0.5780	0.6829
p-value	0.098	0.270	0.364

Research question 2.3f investigates whether there is a relationship between tolerance for ambiguity and the degree of agreement with the information access sequences and average information attention duration. Subjects with a low tolerance for ambiguity would be expected to show more agreement than those with a high tolerance for ambiguity, because they use some kind of benchmark (e.g. the audit manual or the opinion of colleagues) when making decisions and searching for information. Therefore a one-tailed test was applied. The results are summarized in table 23.

It was found that the subjects with a low tolerance for ambiguity showed significantly more agreement with the average information access sequence (p-

value=0.015) and significantly less agreement with the standard information access sequence (p-value=0.053) than subjects with a high tolerance for ambiguity. This finding is understandable in the light of the negative correlation between the standard and average information access sequences. What it tells us is that, of the two ways in which the subjects with a low tolerance for ambiguity could reduce ambiguity, by working methodically through the information in the sequence inherent in the software (so as to miss nothing) or by following a sequence learnt in practice, they tended to choose the latter. No significant differences for information attention duration were found, but the direction of the differences regarding information attention duration is in accordance with expectations.

**Table 23:** Results of independent t-tests on the relationship between degree of agreement with the information access sequences and average information attention duration and tolerance for ambiguity.

Tolerance for ambiguity	Agreement		
	Standard sequence (SS) (1-tail sign.)	Average sequence (CS) (1-tail sign.)	Attention duration (AD) (1-tail sign.)
Low tolerance for ambiguity	-0.1271	0.6705	0.6779
High tolerance for ambiguity	0.1026	0.4478	0.6056
p-value	0.053	0.015	0.165

### 6.3 Summary and conclusions

The objective of this chapter was to determine whether there are relationships between auditors' psychological profiles and demographic characteristics and their problem-solving behavior. Auditors' psychological profiles were considered to consist of: cognitive style for information acquisition, cognitive style for information processing and tolerance for ambiguity. Auditors' demographics were considered to consist of: number of years of auditing experience, educational background and firm affiliation. Auditors' problem-solving was represented by three main categories of variables: operators and task encodings (variables: *IS*, *IR*, *J*, *D*, and additionally *RR*, *TT*), recall and completeness of information access (variables: *RECALL*, *COMP*), and information access sequence and attention (variables: *SS*, *CS*, *AD*).

As regards the relationship between psychological profiles and demographics on the one hand and operators and task encodings on the other hand, it was found that the more experienced subjects applied less information search operators and more

information retrieval operators. The subjects with a NIVRA education applied more information search operators than the subjects with an academic education and so did the sensing types when compared to the intuitive types. However, as has been found in chapter 5, there was no significant observable relationship between information acquisition style and educational background. So the distinction between NIVRA and university training courses does not explain the differences between sensing and intuitive types, or *vice versa*.<sup>4</sup>

Some differences were found between the accounting firms. These differences can best be summarized by indicating the type of operator that is more frequently applied in each firm, as compared to the other firms. We found that firm 1 affiliates are typically information search-oriented, that firm 2 affiliates are typically decision-oriented, and that firm 3 affiliates are typically judgment-oriented. This might indicate differences between accounting firms in organizational culture or personnel selection methods.

Thinking information processing types appeared to have higher relevance ratios than feeling types. This indicates that thinking types need less words to communicate their messages effectively. However, they also needed more time to solve the case, but they applied fewer information retrieval operators.

As expected - considering the finding that most subjects were highly tolerant for ambiguity - no significant differences were found between tolerance for ambiguity categories regarding operators and task encodings.

The findings regarding the relationships between demographics and psychological profiles and operators and task encodings are summarized in table 24.

As regards the relationship between psychological profiles and demographics on the one hand, and operators and recall and completeness of information access on the other hand, an interaction effect between experience and completeness of information access, but no main effects, were found. This is interesting since it indicates that the more experienced subjects are better able to handle large amounts of information (as represented by amount of recall in conjunction with completeness of information access) than the less experienced subjects. Prior research revealed that more experienced subjects could recall more than less experienced subjects [Weber, 1980; Boshuizen, 1989]. Previous studies had not made any allowance for the differing amounts of information which subjects had accessed, so our finding adds a new dimension to the relationship between experience and amount of recall: experience as such does not lead to any significant improved performance in recall tasks in general, but it does have a more positive effect on recall as information exposure increases.

**Table 24:** Summary of significant findings regarding the relationships between demographics and psychological profiles and operators and task encodings.

Research question	Test	Yes/No $\alpha=0.10$	Findings (only significant differences are reported)		
<i>Q2.1a</i> <i>Is there a relationship between the number of years auditing experience and the relative number of operators in each category IS, IR, J and D?</i>	t-test	Yes	Inexperienced IS/NRS=0.11 IR/NRS=0.48	Experienced IS/NRS=0.06 IR/NRS=0.58	
<i>Q2.1b</i> <i>Is there a relationship between the educational background and the relative number of operators in each category IS, IR, J and D?</i>	t-test	Yes	NIVRA IS/NRS=0.11	Academic IS/NRS=0.05	
<i>Q2.1c</i> <i>Is there a relationship between firm affiliation and the relative number of operators in each category IS, IR, J and D?</i>	t-test	Yes	FIRM=1 IR/NRS=0.63 J/NRS=0.28 D/NRS=0.01	FIRM=2 D/NRS=0.03	FIRM=3 IR/NRS=0.50 J/NRS=0.41 D/NRS=0.01
<i>Q2.1d</i> <i>Do sensing types verbalize more information search operators than intuitive types?</i>	t-test	Yes	Intuition IS/NRS=0.07	Sensing IS/NRS=0.13	
<i>Q2.1e</i> <i>Do thinking types verbalize more relevant statements and IR operators and do they take more time for decision-making than feeling types?</i>	t-test	Yes	Thinking RR=0.23 TT=3468 IR/NRS=0.43	Feeling RR=0.18 TT=2734 IR/NRS=0.55	
<i>Q2.1f1</i> <i>Do those with low tolerance for ambiguity verbalize more J and D operators than those with a high tolerance for ambiguity and do those with high tolerance for ambiguity verbalize more IS and IR operators than those with low tolerance for ambiguity?</i>	t-test	No			
<i>Q2.1f2</i> <i>Do those with a high tolerance for ambiguity verbalize more operators from which it is unclear whether the investigation or the audit task is being addressed, as compared to those with a low tolerance for ambiguity?</i>	t-test	No			

As expected, no differences - as to amount of recall - were found between auditors with a NIVRA and an academic education, or between auditors from different accounting firms.

Sensing types recalled significantly more than intuitive types, independent of the amount of information that was accessed. No significant relationship was found between information processing styles - feeling or thinking - and amount of recall, whether in combination with amount of information access or not. Finally, an unexpected significant result was that subjects with a high tolerance for ambiguity recalled more than subjects with a low tolerance for ambiguity, independent of the amount of information that was accessed.

Table 25 summarizes the findings regarding the relationships between demographics and psychological profiles, and recall and completeness of information access.

**Table 25:** Summary of significant findings regarding the relationships between demographics and psychological profiles and recall and completeness of information access.

Research question	Test	Yes/No $\alpha=0.10$		Findings (only significant differences are reported)
<i>Q2.2a</i> Is there a relationship between the number of years auditing experience and the number of recalled relevant statements after correction for information access?	ANOVA	Yes	F-value=7.37	Interaction effect of experience and completeness.
<i>Q2.2b</i> Is there a relationship between the educational background and the number of recalled relevant statements after correction for information access?	ANOVA	No		
<i>Q2.2c</i> Is there a relationship between firm affiliation and the number of recalled relevant statements after correction for information access?	ANOVA	No		
<i>Q2.2d</i> Is there a relationship between information acquisition styles and the number of recalled relevant statements after correction for information access?	ANOVA	Yes	F-value=3.87	Main effect of information acquisition style. Recall sensing > recall intuitive types.
<i>Q2.2e</i> Is there a relationship between information processing styles and the number of recalled relevant statements after correction for information access?	ANOVA	No		
<i>Q2.2f</i> Is there a relationship between tolerance for ambiguity and the number of recalled relevant statements after correction for information access?	ANOVA	Yes	F-value=3.58	Main effect of tolerance for ambiguity. Recall high tolerance > recall low tolerance types.

As regards the relationship between psychological profiles/demographics and agreement with the information access sequences and average attention duration, it was found that relatively inexperienced subjects agreed more with the average attention duration per accessible information page than the more experienced subjects. Educational background and firm affiliation did not lead to differences in agreement with the information access sequences or average attention duration. However, differences in psychological profiles appeared to explain differences in information access and attention behavior. As expected, subjects with a sensing style of information acquisition agreed more with the average access sequence than subjects with an intuitive information acquisition style. Also as expected, subjects with a thinking information processing style agreed more with the standard information access sequence than subjects with a feeling information processing

style. However, agreement with the standard information access sequence was not very high, even within the group of thinking types. Finally, a significant difference regarding information access sequence was found between subjects with a high and a low tolerance for ambiguity. Subjects with a high tolerance for ambiguity agreed more with the standard access sequence and less with the average access sequence than subjects with a low tolerance for ambiguity.

In table 26, the findings regarding the relationships between demographics and psychological profiles and agreement on information access sequence and attention are summarized.

**Table 26:** Summary of significant findings regarding the relationships between demographics and psychological profiles and agreement with the information access sequences and average attention duration.

Research question	Test	Yes/No $\alpha = 0.10$	Findings (only significant differences are reported)	
<i>Q2.3a</i> <i>Is there a relationship between the number of years auditing experience and the degree of agreement with the access sequences or average attention duration?</i>	t-test	Yes	Inexperienced AD=0.72	Experienced AD=0.55
<i>Q2.3b</i> <i>Is there a relationship between the educational background and the degree of agreement with the access sequences or average attention duration?</i>	t-test	No		
<i>Q2.3c</i> <i>Is there a relationship between firm affiliation and the degree of agreement with the access sequences or average attention duration?</i>	t-test	No		
<i>Q2.3d</i> <i>Is there a relationship between information acquisition styles and the degree of agreement with the access sequences or average attention duration?</i>	t-test	Yes	Intuition CS=0.38	Sensing CS=0.66
<i>Q2.3e</i> <i>Is there a relationship between information processing styles and the degree of agreement with the access sequences or average attention duration?</i>	t-test	Yes	Thinking SS=0.05	Feeling SS=-0.08
<i>Q2.3f</i> <i>Is there a relationship between tolerance for ambiguity and the degree of agreement with the access sequences or average attention duration?</i>	t-test	Yes	Low tolerance for ambiguity SS=-0.13 CS=0.67	High tolerance for ambiguity SS=0.10 CS=0.45



## Notes

1. A boxplot is a graphical representation of the central tendency and the variability of data. A box represents the median (horizontal line) and the distance between the 25<sup>th</sup> and 75<sup>th</sup> percentile (the box-length). The vertical lines above and beneath a box represent those values which are outside the 25<sup>th</sup> and 75<sup>th</sup> percentiles, but which are still less than 1.5 box-lengths from these percentiles. Outliers are considered to be between 1.5 and 3 box-lengths from the 25<sup>th</sup> and 75<sup>th</sup> percentiles and are represented by an open dot. Extremes are considered to be above 3 box-lengths from the 25<sup>th</sup> and 75<sup>th</sup> percentiles and are represented by a star.
2. Fisher's Z transformation is aimed at developing an approximately normally distributed variable. The statistic Z is calculated as follows:  $Z = 1/2 \ln \{(1 + R_s)/(1 - R_s)\}$ .
3. Since each financial statement in the case description contains both the requested year figures and the preceding year figures, the trend in income can be determined by accessing only two instead of four Profit and Loss pages.
4. A chi-square test on the association between *EDU* and *ACQ* revealed a non-significant relationship (p-value=0.6958).



---

## Auditors' knowledge representation at the knowledge level

---

### 7.1 Introduction

The main purpose of this chapter is to describe auditors' materiality knowledge at the knowledge level. Three interacting perspectives are to be considered: tasks, models, and methods. Planning stage materiality is considered to influence the nature, extent, and timing of audit procedures [SAS no. 47]. Within the context of the general audit judgment model as presented in chapter 1 of this dissertation, materiality decisions are expected to improve as more expertise is gained in this field. In the preceding chapters of this dissertation we have investigated what personal characteristics might influence auditors' decision-making. It was found that problem-solving behavior varied between subjects and that demographics and psychological profiles might help explain this variability. In this chapter an inventory is made of materiality determining factors. Unfortunately, the materiality problem does not have an unambiguous solution, so that it is not possible to determine what materiality decision is correct. Bernstein [1967] emphasized that subjects' approach to the materiality decision seems to be highly personal, since the decisions can vary significantly based on the same or similar sets of facts. Bonner & Lewis [1990] and Libby & Frederick [1990] stress the importance of differences in innate cognitive abilities between auditors. Especially where the same facts are presented, these differences have great explanatory power with respect to materiality decisions. Given the results obtained in the preceding chapters, different sets of materiality determining variables are to be expected. In this chapter, an overview is given of all the factors that were considered to influence the materiality threshold, regardless of the personal characteristics of the subjects who mentioned each specific factor. So, the general research question investigated here is:

*What factors determine audit planning stage materiality thresholds?*

Section 7.2 provides a brief introduction to the levels of knowledge representation which are generally distinguished in the literature on expert systems development, in order to justify the approach that is followed in the following section. Section 7.3 discusses the results of the research question which is investigated in this chapter. Finally, in section 7.4, the findings are summarized and some conclusions are drawn.

## 7.2. Knowledge representation

Knowledge can be described at the knowledge, symbolic, and physical levels [Steels, 1992, p. 109]. For reasons explained in this section, the knowledge level is especially relevant when describing expertise.

At the knowledge level,<sup>1</sup> knowledge is described at a highly abstract level, in an implementation-independent manner, in order to fill in the gap between task characteristics and human solutions to problems. Three perspectives are relevant: tasks, models, and methods. A task is an interrelated set of activities which is performed by a problem solver and which is aimed at answering the question: 'what has to be done in order to solve the problem on hand?'. Tasks are graphically represented in task structures. A model is a conceptual representation of reality and is aimed at answering the question: 'what knowledge is available in order to solve the problem on hand?'. Models are graphically represented in semantic networks. A method is a set of procedures for the organization and execution of modelling activities and is aimed at answering the question: 'how and when is knowledge used?'. Methods are graphically represented in flow control diagrams.

At the symbolic level, knowledge is described in such a way that it can be implemented into the knowledge base of an expert system, i.e., at a lower abstraction level than when it is described at the knowledge level. Here, a description in terms of rules, objects, facts and inference structures is made. When trying to elicit knowledge on this level, some practical problems emerge. Firstly, experts will normally not communicate their knowledge in terms of if-then statements. Secondly, expert systems which are entirely based on knowledge descriptions on the symbolic level have maintenance problems because of the lack of modularity. Thirdly, rules are often incomplete representations of human knowledge, since there may be underlying knowledge which is not visualized within production rules. For these reasons, a description of human knowledge should never be made entirely at a symbolic level.

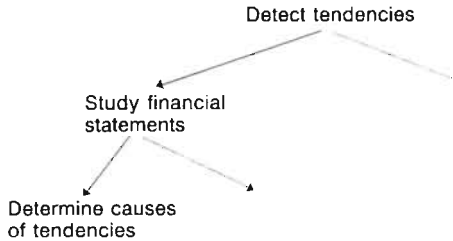
At the physical level, knowledge is described in terms of hardware structures and hardware operations. This is the least abstract representation. The relationship with human problem-solving is not very close, since humans do not think in machine-code or other hardware-imposed symbols. For that reason, a description of human knowledge at the physical level is not relevant when making a representation, as is done in this dissertation, of human decision processes. Furthermore, since the goal of this dissertation is to describe auditors' decision processes in relation to personal characteristics, the development of the materiality expert system itself is not described.<sup>2</sup> There is therefore no description of auditors' materiality knowledge at the physical or symbolic levels.

### 7.3 Results

An analysis of the verbal protocol transcripts acquired in the first round of sessions of the research project yielded information with respect to the tasks which are performed when trying to determine materiality thresholds. This information was not very detailed. For example, one subject concurrently verbalized the following:

*'OK, we are talking about the materiality threshold for the balance sheet, the quantitative materiality threshold. I think that I will start with reading the profit and loss accounts and balance sheets. Subsequently, I will try to find out what tendencies appear to exist within the figures: developments of income and balance sheet. Especially, the relationship between results and balance sheet components, or, in other words: are fluctuations in balance sheet components entirely due to operational activities or are they also a result of investment or finance activities, etc.. So, we will start by forming an opinion with respect to the profit and loss account and the level of business activities.'*

Figure 1: partial task structure as derived from one VPA transcript.



This verbal protocol transcript provides hardly any information regarding materiality or risk determining factors.<sup>3</sup> However, it gives some pointers towards a conceptual description of auditors' decision processes in materiality assessment, in the form of information on the tasks to perform and the methods to be applied in combining tasks. From this transcript, it can be concluded that auditors study financial statements in order to detect tendencies in income and balance sheet components. Furthermore, they are interested in the causes of tendencies. In a task structure diagram, this finding would be graphically represented as shown in figure 1.

By combining the VPA transcripts (concurrent as well as recall), the audit textbook and audit manual knowledge and the additional information acquired in the refinement stage into comprehensive task structures, semantic networks, and flow diagrams, a description at the knowledge level is constructed. Appendices K1, K2,

and K3, respectively, contain graphical representations of:

- the task structure of auditors' decision processes in materiality (and related risk) assessment processes,
- the semantic network of auditors' decision processes in materiality and related risk assessment processes, and
- the logical sequence of auditors' decision processes in materiality assessment processes.

Together, these comprise a description of auditors' decision processes at the knowledge level.

The auditors could be divided into two main groups: those who treat materiality and risk interchangeably, and those who observe the theoretical difference between materiality and risk, and hence cannot accept any trade-off between these concepts. If the technical difference between materiality and risk were followed rigidly the number of rules in the prototype expert system would be reduced by more than 50%.<sup>4</sup>

Based on the technical difference between materiality and risk - materiality is concerned with accuracy whereas risk is concerned with reliability - [e.g. Leslie, 1985; Arens & Loebbecke, 1991; Pany & Whittington, 1994] planning stage materiality should be dependent on the following factors:

- size of the firm, expressed by means of turnover, net income, balance sheet total, or equity,
- shareholders or other participants legitimately requiring an audit with a certain precision,
- application of certain more or less subjective measurement principles in the financial statements (e.g. current versus historical cost),
- whether the company listed on the stock exchange (determines the number of stakeholders),
- the possibility of errors which could not be corrected in subsequent periods (e.g. firms who receive large state subsidies based on financial statements),
- number of subjectively determinable accounts (e.g. provisions for bad debts or for obsolete stocks),
- trend and level of net income (e.g. income approaching to zero leads to lower materiality thresholds),
- contractual agreements which implicitly define the magnitude of the materiality threshold (e.g. in a takeover situation or when negotiating a finance contract),
- type of firm (determines principal financial statement, and potentially the materiality base),
- financial structure (potentially determining the materiality base).

The remaining factors are determinants of audit risk and hence have no influence on materiality.

Appendix K1 shows the task structures of auditors' materiality threshold

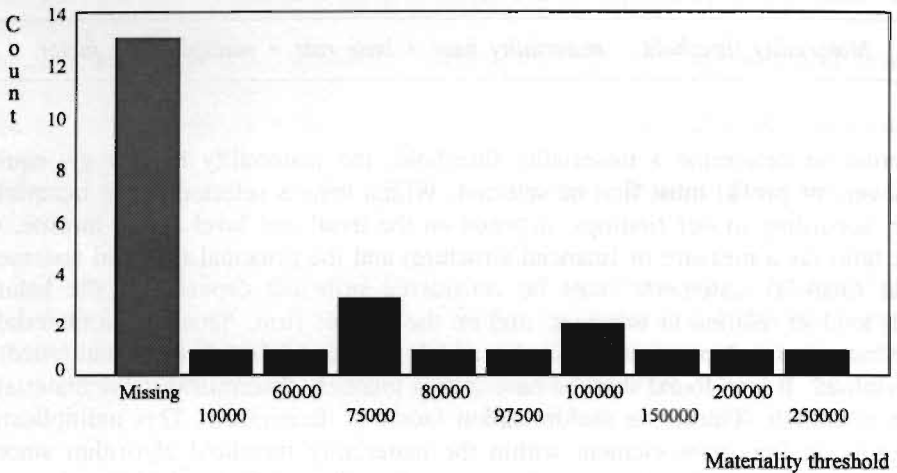
determination processes. The basic algorithm for determining materiality is:

$$\text{Materiality threshold} = \text{materiality base} * \text{base rate} * \text{multiplication factor}$$

In order to determine a materiality threshold, the materiality base (e.g., equity, turnover, or profit) must first be selected. Which item is selected as the materiality base, according to our findings, depends on the trend and level of net income, the debt ratio (as a measure of financial structure) and the principal financial statement. What financial statements must be considered principal depends on the balance sheet total in relation to turnover, and on the type of firm. Secondly, a materiality base rate (a fixed percentage rate by which the materiality base is multiplied) is determined. It was found that the base rate is uniquely determined if the materiality base is known. Thirdly, a multiplication factor is determined. This multiplication factor is an important element within the materiality threshold algorithm since it incorporates qualitative factors into the model. The multiplication factor can be considered to be dependent on audit risk factors and on true materiality determining factors. If the multiplication factor is considered to be dependent on audit risk factors, a division is first made into inherent risks and internal control risks. Inherent risks are assessed on the basis of: tendencies, type of firm, financial accounting and audit environment, and historical evidence regarding the audit and (positive or negative) experience with this client. Internal control risks are assessed on the basis of the quality of the internal organization and the supervisory controls. Within the inherent risk assessment, tendencies - defined as propensity to overstate or to understate - must be determined per segment or cycle and must be seen in relation to breaches in the income trend. Furthermore, the causes of tendencies must be investigated in order to assess the magnitude of inherent risk. If the multiplication factor is considered to be dependent on true materiality determining factors, then the following aspects must be taken into account: influence of shareholders who require a more precise audit, measurement principles (historical versus current), whether the company is listed on the stock exchange, the existence of errors which cannot be corrected in subsequent periods, the size of provisions, the trend and the level of net income, and whether or not materiality is dictated via contracting (e.g. a government might only grant a subsidy if an audit is conducted within a certain range of accuracy).

The semantic network defines the role that each piece of information plays in the determination of materiality. For example, the materiality base and base rate are linked to each other, and the materiality base is defined by three factors: the debt ratio, principal financial statement, and income, which in their turn are defined by more elementary factors (see appendix K2).

The sequence of each task within the process of materiality determination is shown

**Figure 2:** Bar diagram of materiality threshold frequencies.

in appendix K3. The main process is that of determining the materiality base, base rate and multiplication factor. The process of determining the materiality base and hence the base rate is described in the diagram below. Each time a sub-process cannot be completed, a 'fail' is encountered. Only when a sub-process is completed may the next sub-process be initiated. The magnitude of the correction factor is determined through an iterative process, represented by the arrows in the diagram in the third part of appendix K3. Three sub-processes play a role in determining the multiplication factor: analytical review, understanding and recording the system, and gathering evidence as to the control environment.

During the protocol sessions and the experimental testing of the expert system prototype it became apparent that materiality - as already noticed by Bernstein [1967] - is a very subjective concept. Based on the very same case, materiality thresholds varying from 10,000 to 250,000 (mean 106,042) were determined by 12 subjects (the remaining 13 did not communicate any materiality threshold). Figure 2 is a bar diagram of these thresholds.

#### 7.4 Summary and conclusions

This chapter has followed an intuitive approach to the process of knowledge elicitation from expert auditors regarding materiality. Auditors' materiality assessment knowledge has been acquired from various sources:

1. verbal protocols (concurrent and recall);
2. textbooks and audit manuals;
3. prototype expert system refinement sessions.

Each materiality determining factor that was mentioned by at least one subject in



the VPA and system refinement sessions was recorded and analyzed in order to develop an insight into task structures, models and methods as applied in the process of materiality determination. It should be stressed that no rigid research methodology has been followed in gathering the data as represented in this chapter. Rather, we have tried to gain an overall view of materiality determining factors. Hence, this approach is primarily goal-oriented. The findings from this chapter are summarized in table 1.

Table 1: Summary of findings regarding materiality determining factors.

Dependent variables		Determiners	
Materiality threshold		materiality base * base rate * multiplication factor	
- Mean = 106,042			
- Min = 10,000			
- Max = 250,000			
Materiality base	- trend and level of net income		
	- debt ratio		
	- principal financial statement		
Base rate	- materiality base		
Multiplication factor	- audit risks	- tendencies to overstate or to understate	
		- type of firm	
		- audit holding company	
		- term accounts ready	
		- audit this client (tenure)	
		- management fraud	
		- quality of internal organization	
		- supervision	
	- true materiality determining factors	- influence of shareholders requiring an audit with a certain precision	
		- subjectivity of measurement principles	
		- company listed on the stock exchange	
		- errors which cannot be corrected in subsequent periods	
		- size of provisions	
		- trend and level of net income	
		- materiality threshold is dictated via contracting	

Further research is needed on the relationship between what is considered material or materiality determining, on the one hand, and personal characteristics

(demographics and psychological profiles) on the other hand. Here, it should be borne in mind that expertise in this field might be non-existent, because of the unverifiable nature of materiality decisions and the resulting lack of feedback and learning possibilities. Eventually, this research might reveal different sets of materiality determining factors for different types of persons. For now, the description at the knowledge level, as given in this chapter, might be a good starting point for further research that uses more rigid research methodologies and which takes account of differences in demographic characteristics and psychological profiles.

## Notes

1. Although it might be confusing to describe a concept at a level that has the same name as the concept itself, we use the wording: '*knowledge level*' for describing expert *knowledge* at a highly abstract level because the leading textbooks on this topic use this terminology (see e.g. Hickman *et al.* [1989], Steels *et al.* [1992]).
2. For a description of the prototype expert system that was developed in this study, see Vaassen [1991].
3. Encoded into relevant statements and operators or task categories, this part of the literal protocol of subject number 6 reads as follows:
  - If the goal is to determine a quantitative materiality threshold then I will start by reading the profit and loss accounts and the balance sheets (coded: IS/I).
  - I will try to find out what tendencies appear to exist within the income statement and the balance sheet (coded: IS/I).
  - Especially important is whether fluctuations in balance sheet components are entirely due to operational activities or are also due to investment or finance activities (coded: IS/I).
4. From 85 rules to just 42, the number of rules referring to true materiality determining factors.



---

## Summary and discussion

---

### 8.1 Summary

#### 8.1.1 Summary of research methodology

The main research questions addressed in this study are:

1. Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?
2. Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?
3. What factors determine audit planning stage materiality thresholds?

In order to address these questions, 25 experienced auditors from four international public accounting firms were interviewed by means of verbal protocol analysis techniques, personality questionnaires and expert system refinement techniques.

In the first round of sessions, subjects were presented with a case displayed by a computer program whose automated process tracing function kept track of information page access sequences and information page attention durations. The subjects were asked to think aloud while solving the case. All the verbalizations were recorded on audio and video tape and then transcribed literally. After having solved the case, the subjects were asked to fill out two questionnaires: a Myers-Briggs information acquisition and processing questionnaire, and a Tolerance for Ambiguity questionnaire. Furthermore, some demographic information was recorded: their number of years of auditing experience, whether they had had their formal education via the institute of registered accountants (NIVRA) or via one of the universities, and their firm affiliation.

The data was analyzed in four steps: (1) the literal protocol transcriptions were subdivided into relevant statements, (2) the relevant statements were scrutinized in order to extract if-then rules for the development of a prototype expert system to be used as an instrument for gathering data on materiality determining factors, (3) the relevant statements were encoded into four categories of operators and four categories of tasks, and (4) the personality test scores and the demographic factors were analyzed and related to the automated process tracings and to the verbal protocol encodings.

In the second round of sessions, a prototype expert system was presented to 17 of 25 subjects to gather additional data on materiality determining factors and to refine the prototype expert system. The refined prototype expert system was then presented to 5 (of the remaining 8) subjects, in order to examine its performance and to gather more data on materiality determining factors.

### **8.1.2 Summary of results**

An overview of the research questions and their interrelationships is given in appendix G. This section summarizes the results for each of the main research questions.

#### **Results concerning psychological profiles (Q1)**

Auditors appeared to be generally highly tolerant to ambiguity, but could not be classified into any particular information acquisition (sensing versus intuition) or processing (feeling versus thinking) category. There appeared to be no significant relationship between experience and educational background on the one hand and psychological profiles on the other hand. However, there appeared to be differences between auditing firms as to the tolerance for ambiguity and information processing styles of their affiliates: auditors from firm 1 had a preference for feeling type information processing and were more tolerant to ambiguity than auditors from firm 2 and - although to a lesser degree - than auditors from firm 3. Since, firm 1 is the least structured firm within our study, this finding is in accordance with expectations.

#### **Results concerning psychological profiles in relation to problem-solving behavior (Q2)**

In terms of the general audit judgment model as presented in chapter 1, some interesting results due to differences in demographic characteristics were found. It appeared that experienced auditors were less information search oriented and more information retrieval oriented. Also, NIVRA-educated auditors were more information search oriented. Furthermore, auditors from firm 1 were more information retrieval and less judgment oriented than auditors from firm 3, auditors from firm 2 were more decision oriented than auditors from firms 1 and 3.

Regarding psychological profiles, some interesting relationships were found. Sensing information acquisition types appeared to be more information search oriented than intuitive types. Feeling types appeared to need more words to communicate their messages effectively, but took less time to solve the case than thinking types. These findings are in accordance with expectations. However, it was also found that feeling types were more information retrieval oriented. This finding is not in accordance with expectations. No significant relationship was found between tolerance for ambiguity and problem-solving behavior, as measured

by operator frequencies or task codings. This is in accordance with expectations since most subjects were highly tolerant to ambiguity.

As to the relationship between demographic characteristics and amount of recall, it was found that the more experienced subjects were able to recall more information when they had accessed more information during problem-solving. This is an expected finding. However, the less experienced subjects, surprisingly, could recall less information if they had accessed more during problem-solving. It could be that there is an information overload effect when the subject has less experience in handling large amounts of information. No differences regarding education or firm affiliation were found.

With respect to psychological profiles, it was found that sensing types recalled significantly more than intuitive types, independent of the amount of information that was accessed during problem-solving. Furthermore, subjects with a high tolerance for ambiguity recalled more than subjects with a low tolerance for ambiguity, independent of the amount of accessed information. No differences regarding information processing style and amount of recall were found.

Regarding inter-subject agreement on information access sequence and attention duration, it was found that 3 subjects showed a significant ( $\alpha=0.10$ ) positive correlation and 1 subject showed a significant negative correlation with the standard information access sequence, while 18 subjects showed a significant ( $\alpha=0.10$ ) positive correlation with the average information access sequence, and 22 subjects showed a significant ( $\alpha=0.10$ ) positive correlation with the average information attention duration.

When we related differences in agreement with the information access sequences and average information attention duration to differences in demographic characteristics, it was found that the more experienced subjects agreed less with the average durations for each information page. No other differences as a result of different demographic characteristics were found.

When we related differences in agreement with the information access sequences and average information attention duration to differences in psychological profiles, it was found that sensing information acquisition types agreed more with the average information access sequence than intuitive types, that thinking types agreed more with the standard information access sequence than feeling types, and that subjects with a high tolerance for ambiguity agreed more with the standard information access sequence and less with the average information access sequence than subjects with a low tolerance for ambiguity.

### **Results concerning materiality determining factors (Q3)**

Auditors appeared to exhibit no consensus as to the magnitude of the materiality threshold in the case used in this study. Based on this finding it may be assumed that auditors perceive information differently and hence process that information differently. As a result, materiality thresholds may vary considerably. However, some general procedures and information components which may be further investigated as to their relevance in materiality determination were identified.

It was found that materiality is primarily determined on the basis of quantitative indicators such as turnover, balance sheet total, equity and income in combination with fixed percentage rates. However, the materiality thresholds which are at first determined may be corrected in the light of circumstances that require materiality thresholds to be higher or lower. These specific circumstances are qualitative, and are the most difficult elements in the materiality determination process.

The basic algorithm for materiality threshold determination is:

materiality base \* base rate \* multiplication factor. The materiality base and base rate are typically quantitative elements, whereas the multiplication factor is typically a qualitative element within the algorithm. Some factors that determine the materiality base and hence the base rate are: the trend and level of net income, the debt ratio, and the principal financial statement. The factors that determine the multiplication factor can be sub-divided into audit risk factors and true materiality determining factors. Audit risk factors are often mentioned as materiality determining factors, although audit risk is a measure of reliability whereas materiality is a measure of accuracy. Given the relationship between materiality and risk (see chapter 1) in practice, the consequences of lowering the materiality threshold are similar to the consequences of increasing the risk estimate. This makes the implications of erroneously considering risk factors as materiality determining factors considerably less serious.

## **8.2 Discussion**

In this final section, the practical and theoretical implications of the current study are discussed. It appears that some limitations must be borne in mind when interpreting the results of this study. Yet these limitations have a positive effect, since they can generate interesting research proposals. This dissertation concludes with a number of such research proposals.

### **8.2.1 Practical implications**

What results have been achieved in relation to the needs of audit practice? Most of the earlier audit judgment research has had no clear-cut implications for practice. But the decision aid studies conducted by Weber [1978], Cushing & Loebbecke



[1986], and Murphy [1990], in particular, do have practical implications. These provide evidence on the degree of structuring which is desirable in audit manuals or for the use of expert systems in auditing. Expert system studies, which are mainly funded by auditing practices, provide practical results almost by definition, since they are aimed at the development of a decision aid which is intended to be used in practice [e.g. Hanssen & Messier, 1986; Dillard & Mutchler, 1986]. Most audit judgment studies, apart from decision aid research, have only an indirect relevance for practice, in the sense that a good description of the observed behavior of practitioners could contribute to the development of a normative framework to be used by those practitioners.

A better understanding of the nature of audit expertise will presumably contribute to the improvement of educational programs at universities and auditing courses and to the in-house training courses of public accounting firms. One potential result of improving auditors' formal education would be to reduce the time novices need to become experts. Since individuals differ as regards a number of personal factors which are usually innate, and adjustable to only a limited extent, 'pencil and paper' personality tests might be helpful in determining novices' potential for becoming auditing experts. These personality tests might also be helpful in determining what educational programs would be most effective, taking subjective cognitive styles (as part of personality) into account. Finally, information provided to the auditors associated with a firm might be customized to their specific requirements, which are partially based on their personalities.

Consensus measurement can be considered one of the central themes in audit judgment research as it has been conducted in recent decades. This is because sound competition among public accounting firms should ideally be based on the marketing instrument 'product' - which is what there should be consensus about - and not on 'price', 'geographical', or 'promotional' differences. However product quality is difficult to measure objectively because audit decisions are based on choices out of many options, all of which might be correct, given certain assumptions. The inter-subjective measure of consensus has evolved as a substitute measure of quality. Better understanding of the degree of consensus and its implications for the nature, extent and timing of audit procedures might increase the acceptance and hence application of auditing procedures prescribed by the firm. Moreover, worldwide application of basically identical audit procedures contributes to the high quality image which public accounting firms strive for.

An example of research into the application of such basically identical audit procedures is research with respect to auditors' ability to quantify prior probabilities (to be seen in relation to risk assessments according to an audit risk model). This research has revealed that there is only a weak regression to an average base-rate [Winkler, 1967; Joyce & Biddle, 1981b]. This could be an

indication that audit approaches prescribing an exact risk quantification could lead to great differences in risk assessment. It could also be that the elicitation method or the way the question about risk quantification is asked have an influence. Corless [1972] and Felix [1976] addressed this problem. Both studies concluded (the former more strongly than the latter) that the elicitation method has some influence on auditors' probability estimates. These findings provide a basis for useful practical recommendations, since audit manuals can be so structured that the audit approach is communicated in an optimal way.

Another example of research into the application of basically identical audit procedures is the research into degree of structure of public accounting firms. Cushing & Loebbecke [1986] describe audit firms in terms of degree of structure, whereas Bamber & Snowball [1988] investigated consensus differences between the associates of structured and unstructured firms. Degree of structure appeared not to influence consensus. Joyce & Biddle [1981a] investigated the effects of the anchoring and adjustment heuristic. Auditors putting too much weight on 'anchors' and working in firms with a structured audit approach will presumably make worse materiality estimations. The practical implications lie in the recommendations to be made to auditing firms: to structure or not to structure.

Deductions such as those in the two examples above must be made with caution. In order to conduct research that has major practical implications, a framework for audit judgment research should be developed which focuses primarily on specifying all the relevant relationships between earlier research findings and audit practice. Such a framework would contribute to the development of a coherent set of empirical research findings which provide both a thorough description of audit practice and a basis for making recommendations for audit practice. The general audit judgment model described in the first chapter of this dissertation is an attempt in that direction.

### **8.2.2 Theoretical implications**

It should be possible to construct an audit judgment theory based on audit judgment research as it has been conducted in recent decades. Such a theory would be well grounded in empirical observations and would substantially contribute to our insight into thought processes and related processes such as data storage in human memory, retrieval from memory, information search and learning by experience, and it would be so constructed as to be useful in practice.

Here, the general audit judgment model described in chapter 1 may serve as a benchmark in the development of a generic audit judgment theory, provided that each variable and each interrelationship within that model is tested empirically. In this dissertation, some parts of the general audit judgment model are investigated

by means of a set of hypotheses. So, a step has been made towards an audit judgment theory.

Some authors have discussed the relationship between psychological profiles - as measured by various personality and cognitive style tests - and actual problem-solving behavior [Driver & Mock, 1975; Kerin & Slocum, 1981; Schweiger & Jago, 1982; Pincus, 1990]. In relation to the concept of expertise, Bonner & Lewis [1990] state:

*'In this study we explore a view of expertise in which specific experiences and training create knowledge, and knowledge is combined with innate ability to perform specific audit tasks.'*

This view of audit expertise is followed in our study: expertise is considered to develop through domain-specific formal education, on-the-job experience, and innate abilities to perform well in specific audit tasks and to learn effectively from experience.

### 8.2.3 Limitations of the research

As has been stipulated in the preceding section, this dissertation is a step towards an audit judgment theory. However, some issues have to be resolved before generalizations can be justified. As in almost every research project, in spite of careful research design, there are limitations which must be borne in mind in interpreting the results obtained in this study. This section gives an overview of these limitations.

Tuning problems can result when a multimethod research approach is employed, and this has in fact occurred. Verbal protocol analysis, for reasons of efficiency, is usually conducted among small groups of subjects ( $<15$ ), whereas personality assessment studies, for reasons of effectiveness, are usually conducted among larger groups ( $>100$ ). The number of subjects in this research is high when seen in relation to the verbal protocol section, but low in relation to personality assessments. This puts an important limitation on the generalizability of the results obtained.

As has been observed, experience effects may not be significant when a certain critical experience level has been reached. This critical experience level is probably somewhere between 1 and 5 years of auditing experience. However, no research has been conducted to determine the exact amount. The subjects in this research are all very experienced, with at least eight years auditing experience. We would therefore not expect significant experience effects. Had the subjects had a greater range of experience levels more significant results might have been achieved.

When applying verbal protocol analysis, the development of a case is the most important part of the experimental design. In hindsight, the Midcom Ltd case as used in this research could have been pre-tested in greater detail. However, it has indeed been pre-tested and it is debatable where pre-testing must stop and experimenting must start. Any experimental case is limited in nature. For reasons of efficiency, only a small number of items of information can be built into an experimental case: the Midcom Ltd case focused on a family-owned production company, which had some internal control problems, and which was involved in a takeover by a larger firm. In terms of audit risk models: inherent risk and control risk are both considered to be high. A more extensive case would have demanded too much time from each test person or would have required more subjects.

In this research there is no control for the effects of thinking aloud on problem-solving behavior. A literature survey [Ericsson & Simon, 1984] reveals that the effects of thinking aloud cannot *a priori* be determined. The lack of any control for these effects may or may not have affected the strength of the results, we simply don't know.

#### 8.2.4 Suggestions for further research

The first field suggested for further research relates to the fact that any questionnaire-based cognitive style test lacks a certain (mundane) realism since it measures cognitive style as if it consists of stable personal characteristics. However, these characteristics are not as stable as would be expected considering the vast amount of research effort in the cognitive styles field, which almost by definition is static research (see chapter 2). For this reason Rosenberg [1968] has proposed a different term for the variables resulting from cognitive style questionnaires: 'dispositions', referring not to an individual's state or condition but to a tendency to react in a certain way under certain circumstances. The so-called stable personal characteristics measured by a personality questionnaire may not show up under dissimilar circumstances. Cognitive style research is based on the assumption that personality as measured by some well-known questionnaires (e.g. the MBTI or the TFAM) should account for individual differences in decision-making. Therefore these questionnaires must be subjected to a validity test. The experimental design would be as follows: construct a large cognitive style questionnaire, with about 100 items (like the original MBTI questionnaire). Have a group of decision-makers fill in this questionnaire, and then change the subjects' knowledge state by having them solve an experimental problem. This problem must be sufficiently time-consuming and difficult to ensure that subjects will have to change their knowledge state. Furthermore, a self-insight test must be built into the case description. Later the original cognitive style questionnaire is presented to the subjects. The experimental results are measured by means of the correlation coefficient between the two questionnaire responses, together with the degree of

self-insight. The hypothesis is that self-insight and the correlation coefficient between the two questionnaire responses will be positively correlated. Confirmation of this hypothesis would be evidence for the rejection of the Rosenberg proposition that cognitive style questionnaires measure only dispositions, not stable personal characteristics.

The relationship, if any, between personality characteristics on the one hand and audit experience on the other hand offers a related research topic. A relationship in any direction would indicate that personality (including cognitive style) is subject to change over time and hence must be seen as a disposition rather than a stable personal characteristic. A population of auditors with varying levels of experience would have to be studied. As was demonstrated in the current research, audit experience beyond 5 years does not lead to differences in personality characteristics or any other differences in information search behavior. So experience must either be measured on another scale or it must be varied from a minimum level of zero to the maximum possible in order to assess any differences. Also, the 'critical experience level' could be assessed by having subjects with experience levels varying from zero to, for example, 40 years perform a task with an unambiguous solution.

The second research field is the relationship between tolerance for ambiguity and decision confidence. This could be investigated by asking a representative sample of experienced auditors what confidence (scale 0 to 10) they would assign to a choice of a certain materiality threshold adjustment on the basis of a set of client-specific circumstances that affect inherent risk. The average score per subject would indicate the degree of decision confidence and can be considered to be a performance-based measure of tolerance for ambiguity. The higher the average decision confidence, the lower the tolerance for ambiguity. Moreover, the questionnaire-based test could be validated by means of this performance-based instrument.

Third, verbal protocol analysis is subjective in nature. Much of its quality depends on the individual skill, effort and knowledge of the encoder. Two methods have been proposed to increase objectivity: the multi-encoder method and the single encoder/elapsed time method. It is also rationally expected that the greater the number of encoding categories, the lower consensus (multi-encoder) or consistency (single encoder) will be, since a greater number of categories implicitly involves more arbitrary allocations. So, an interesting experiment would be to test for consensus and consistency within two encoding schemes (few/many categories) and two approaches to subjectivity (multiple or single encoder).

Fourth, in the current study some personality variables have been investigated in relation to firm affiliation and educational type. Although a sample size of  $n=25$

might be sufficient if it were analyzed integrally, the division of this population into two sub-populations by education or into three groups by firm affiliation results in sample sizes which may be too small. Because the focus in this study is on verbal protocol analysis, the 'small sample experiments' on these two variables serve merely as indicators of tendencies which might be present. Sample sizes must be larger to increase validity. In a follow-up program, the personality questionnaires and a short debriefing questionnaire could be presented to more representatives of the three auditing firms which participated in our project.

Fifth, differences between personality types could be tested by presenting subjects with a broad range of tasks. The results would contribute to our knowledge of differential performance peaking in auditing. Thus far, the results in this field are not conclusive: Benbasat & Dexter, [1979], for example, found evidence for differential performance peaking whereas Lusk [1979] did not. Both studies applied an embedded figures test. However, the task environment can play an important role when investigating interaction effects between personality and performance. Therefore, a model incorporating three variables (performance, personality and task) could be developed and tested using two-way ANOVA.

Sixth, the relationship between materiality and desired audit risk is generally expressed via a convex function. Presumably, therefore, there is a trade-off between the two factors. Research into the exact coefficients of this trade-off function could have some practical implications, since reliability and accuracy in auditing are assessed separately, but are often confused (as shown in this dissertation).

Seventh, much has been said about the determination of the amounts which should be considered material, but the question as to the impact of raising or lowering planning stage materiality thresholds on the nature, extent and timing of audit procedures has hardly been addressed. As has been illustrated, the only quantification of the relationship between planning stage materiality and audit procedures is found in sampling applications, i.e. the extent of audit procedures. The impact on the extent of other substantive audit tests or on the timing of these substantive tests has not been explicitly described. An experiment could be designed as follows: have auditors perform a planning stage audit task in which they are asked to state their agreement or disagreement (on a five-point Likert scale) with certain audit procedures in a well-defined audit environment of limited scope. Three experimental settings would be tested: first, planning stage materiality would be presented as part of the case description, second, subjects would be asked to determine the planning stage materiality as part of the experimental task, and third, it materiality is neither given nor requested. If there are no differences between the three groups this would indicate that, although materiality is said to

have a critical influence on the nature, extent, and timing of audit procedures, in practice it has no influence.

Eighth, some research [Weber, 1978; Abdolmohammadi & Wright, 1987] has revealed that experienced auditors evaluate internal controls to be stronger and calculate smaller sample sizes. However, those experienced auditors developed more elaborate audit programs. These findings could indicate that expert auditors are better equipped to plan audits, a task which they perform as exhaustive as possible, whereas novice auditors focus primarily on certain specific problem domains and try to treat these problems as exhaustively as possible (larger sample sizes, internal controls evaluated to be weaker). In technical terms: semantic memory, which is well-developed among recent graduates, presumably contains thorough knowledge that is limited in scope, whereas episodic memory, well-developed among experts, contains less thorough knowledge that is broad in scope. This hypothesis should be tested, since confirming evidence would imply that accounting firms should develop detailed job descriptions differentiated by experience (or hierarchical) level.

Ninth, as Libby [1982] stipulates, personality assessment may only be relevant for personnel selection. That proposition would point towards more research into the reasons for students dropping out from accountancy study and, more importantly, why recent graduates employed in accounting firms leave the profession. A better understanding of the factors underlying the latter phenomenon would undoubtedly lead to more efficient personnel selection procedures.

Tenth, Bedard [1989], in a literature survey, investigates whether expertise in auditing exists or not. A more specific research topic would be to test whether or not expertise in materiality assessment exists. This research question has not been investigated in this dissertation. It has simply been assumed that the auditors in the sample, given the well-defined selection criteria, were experts in the field of planning an audit and hence in planning stage materiality assessment. However, results indicate that auditors have trouble in quantifying planning stage materiality - only 13 out of 25 auditors quantified a materiality threshold - and in distinguishing between materiality and risk. Moreover, there was hardly any consensus with respect to quantitative planning stage materiality thresholds. So a final question posed by our findings is: 'is planning stage materiality indeed non-existent, or are there serious weaknesses in the research design?'. Considering the results found in this study, we would give priority to this suggestion for further research.

Eleventh, as has been stipulated in chapter 7 of this dissertation, differences in perception of potential materiality determining factors might lead to different materiality assessments in identical or similar cases. These differences in perception of potential materiality determining factors might be explained by differences in the

psychological profiles of subjects. An interesting research field would be to investigate what type of information determines the magnitude of materiality thresholds for various (psychological) types of subject.

Twelfth, the results of our study show (among other things) that there is no significant relationship between experience and psychological profiles. This implies that experience might be investigated in interaction with psychological profiles in order to assess their effects on expertise. Also, in order to investigate the relationship between experience and formal education, an effective research design would have to incorporate both a sample from specialists in a specific audit task (e.g. an analytical review task), non-specialists with respect to that specific audit task and non-specialists with respect to any audit task (e.g. an economist) in order to investigate interaction effects. However, in order to effectively perform two- or three-way ANOVA-procedures, a relatively large sample size is needed. An interesting research proposal would be to distinguish experience-categories, education-categories and psychological profile-categories in an audit task that has an unambiguous solution.

### 8.3 Concluding remarks

When a research project has been completed, one can ask 'what have we learned?'. Currently there is an ongoing discussion about the contributions of empirical accounting and auditing research to the auditor's profession [Bindenga, 1993; Biggs *et al.*, 1994; Van de Poel, forthcoming]. The central theme in this discussion is: 'practical relevance'. In this dissertation we have striven for a research approach that conforms to internationally accepted standards of scientific inquiry. However - and this should always be kept in mind when doing research into human decision processes - if we want to improve decisions that are actually made by practitioners, we will have to understand how these decisions are made. In that sense, the research that has been described in this dissertation has not only theoretical but also practical relevance.



---

## Appendices

---

## Appendix A

### The AICPA audit risk model and the Leslie/CICA audit risk model

#### (1) AICPA (joint risk model)

$$AR = IR * CR * DR$$

#### (2) Leslie/CICA (posterior risk model)

$$AR = \frac{IR * CR * DR}{IR * CR * DR + (1 - IR)}$$

Where:

$AR$  = audit risk

$IR$  = inherent risk

$CR$  = internal control risk

$DR$  = detection risk

The assumptions that underlie each of these risk models can be presented schematically by means of the following diagram (based on Leslie [1985]):

Event	Chance
Material error does not occur	$1 - IR$
Material error is detected by internal controls	$IR \times (1 - CR)$
Material error is detected by the auditor	$IR \times CR \times (1 - DR)$
Material error is not detected by the auditor	$IR \times CR \times DR$

Each of the right-hand side factors marks the combined probability of a material error occurring, given the underlying probabilities ( $IR$ ,  $CR$  and  $DR$ ) at each stage.

In summary:

1. There is a  $1 - IR$  chance of no material error occurring in the first place.
2. There is a  $IR \times (1 - CR)$  chance of a material error occurring but being detected by internal controls.
3. There is a  $IR \times CR \times (1 - DR)$  chance of a material error occurring, not being

detected by internal controls but being detected by the auditor.

4. There is a  $IR \times CR \times DR$  chance of a material error ending up undetected on the financial statements.

The joint risk model determines audit risk as the right-hand bottom of the diagram ( $IR \times CR \times DR$ ) divided by the sum of *all* the right-hand factors (i.e. 1). The posterior risk model, on the other hand determines audit risk as the right-hand bottom of the diagram ( $IR \times CR \times DR$ ), divided by the sum of the right-hand bottom and the right-hand top of the diagram  $(IR \times CR \times DR) + (1 - IR)$ . The reasoning here goes along the following lines. Since there are only two possible right-hand factors when no evidence of a material error is found, the denominator consists of the sum of these two factors, that is:  $IR \times CR \times DR$  and  $1 - IR$ . Hence, the chance that an existing material error is undetected is:

$$AR = \frac{IR * CR * DR}{IR * CR * DR + (1 - IR)}$$

The AICPA model is based on the assumption that all the right-hand factors of the diagram are possible. This is an unrealistic assumption since the audit risk is defined for the situation in which no evidence of a material error has been found.

The audit risk calculated according to the posterior risk model always is greater than or equal to the audit risk calculated using the joint risk model. The difference in outcomes between the two models increases as the prior probability  $IR$  increases.

For this dissertation, the difference between the two models is of minor importance, since, no mathematical relationship between risk components is assumed in any of the analyses performed.

## Appendix B

### The mathematical relationship between materiality and risk

Suppose the materiality threshold for a certain segment is 100 and the total value of that segment is 2,000. The critical error fraction ( $f$ ) then is:  $100/2000 = 0.05$ , indicating that audit procedures are designed such that any error above 100 will be detected. In this situation the following detection risk distribution holds:

- the chance of not detecting a material error in a sample of  $n=1$  is:  $(1-0.05)^1 = 0.95$ ,
- the chance of not detecting a material error in a sample of  $n=2$  is:  $(1-0.05)^2 = 0.903$ ,
- the chance of not detecting a material error in a sample of  $n=3$  is:  $(1-0.05)^3 = 0.857$ ,
- the chance of not detecting a material error in a sample of  $n=150$  is:  $(1-0.05)^{150} = 0.000$ .

Let  $f$  = the critical error fraction, and  $n$  = the sample size, then the expression:

$$(1-f)^n$$

can be transcribed as:

$$(1-f)^{1/f \cdot n \cdot f}$$

Since the critical error fraction approaches zero (usually between 0.01 to 0.05), the component:

$$(1-f)^{1/f}$$

approaches  $1/e$ , where  $e$  = Euler's constant. So, the chance of not detecting a material error in a sample of  $n$  is:

$$(1/e)^{n \cdot f}$$

Let  $MTDR$  = maximum tolerable detection risk, then:

$$\ln(MTDR) = -n \cdot f$$

$$n = \frac{1}{f} * (-\ln(MTDR))$$

$$n = \frac{\text{amount}}{\text{materiality}} * (-\ln(MTDR))$$

Here the relationship between materiality and maximum tolerable detection risk becomes clear:

- the higher materiality, the smaller the sample size,
- the higher maximum tolerable detection risk, the smaller the sample size.

Keeping the sample size fixed, there is a trade-off between materiality and maximum tolerable detection risk. This trade-off is graphically presented by means of a downward sloping curve.

### Audit Judgment Research

In chapter 2 of this dissertation, an overview of audit judgment literature is presented insofar as these studies comply with both of the following criteria:

- there is a focus on materiality and risk judgments in auditing,
- use is made of a dynamic research methodology.

Only a few examples of the remaining audit judgment studies are mentioned. This appendix provides a broader overview of audit judgment literature. The results, grouped by categories, are summarized in a table at the end of this appendix.

#### Dynamic research: decision aids

No findings are reported in the table as regards expert system development, as a part of decision aid research, since the expert system research projects discussed in chapter 2 are aimed mainly at the construction of an expert system<sup>1</sup> rather than at describing audit expertise. The practical relevance of expert systems can easily be proved, but knowledge elicitation and structuring is much more interesting. As with expert system development, the practical relevance of the research to audit standard setting and regulation at the firm level is beyond doubt. The research topic to be addressed here is the effect of decision aids on expertise development.

Mock, Wright, Washington & Krishnamoorthy [1993] determined operators on the basis of verbal protocols. They made a classification of operators, which seems promising for further audit judgment research.

Expert system development, although initiated in the thirties with the work of Turing and Shannon who both, independently,<sup>2</sup> studied the potential for artificial intelligence, did not blossom until the late seventies [e.g. McCarty, 1977]. The pioneering efforts in accounting and auditing were in the field of taxes. Later, when the Peat Marwick Mitchell Research Opportunities Program in Auditing was founded [Gwilliam, 1987], expert systems in other fields of auditing were developed: assessment of adequacy of allowance for bad debt [Dungan, 1983], internal controls in the purchasing cycle [Meservy, 1985], corporate tax accrual and planning [Shpilberg & Graham, 1986], evaluation of internal revenue controls [Gal, 1985], computer audit controls [Hansen & Messier, 1986], internal controls in the sales cycle [Grudnitski, 1986], and auditor's opinion based on going concern

---

<sup>1</sup> These projects are often funded by third parties and hence have a greater incentive to provide results that can be applied in practice.

<sup>2</sup> Turing described the design of a machine that should be able to solve problems the way humans do, recognizing that it would only partially be able to describe human problem-solving (i.e. the logic component). Shannon tried to define electronic circuits in terms of human logic.

judgments [Biggs & Selfridge, 1986; Dillard & Mutchler, 1986].

Swagerman [1989] developed an expert system for auditing expert systems (META). The system classifies expert systems into one of five confidence categories varying from very reliable to unreliable. It has self-learning capabilities in that when pre-classified examples are presented to the system it changes its knowledge base. The categories in META are represented in frames.<sup>3</sup> The system was built in an expert shell (CAKE) that was especially suitable for emulating humans' categorizing.

Vasarhelyi & Halper [1990] developed an expert system for the continuous audit of online systems, designed to deal with the problems of auditing large paperless database systems (working title: CPAS). What happens in this methodology is that transactions which are inputs into the system are monitored and analyzed continuously during processing, using a set of auditor-defined rules. System alarm messages call the auditor's attention to any anomalies in the system. This implies that continuous process auditing is a typical analytical review technique. The methodology has great advantages with regard to the efficiency of the audit. Vasarhelyi & Halper [1990] state: 'Ultimately, if a system is monitored over time, the auditor can audit by exception only'. However, despite this obvious advantage, there is a major technical problem when implementing and applying a continuous audit system: placing probes into a large computer system to monitor its working may reduce its performance with regard to the system's primary objective. The knowledge on which CPAS is based was elicited by interviewing auditors and examining audit working papers, followed by a prototype testing phase in order to discover unformulated rules and enlarge the knowledge base of the system.

Among the main topics in expert system development in auditing are internal controls, taxes, going concern judgments, risk assessments and materiality. Furthermore the development of meta-auditing expert systems is a field with growing potential as auditors have more and more to deal with expert systems as part of the audit environment. Equally, as auditors have increasingly to deal with large databases and paperless administration systems, the need for continuous process audit systems has become obvious. Holstrum, Mock & West [1990] discuss the implications of advancing technology for auditing and internal control. They performed a Delphi study in which 38 experts predicted changes in five areas of business technology, identified by exploratory interviews, for the years 1990 and

---

<sup>3</sup> Waterman [1986] distinguishes three main categories of expert systems: (1) knowledge representation using rules (if-then-statements), (2) knowledge representation using semantic nets (networks of objects, events and concepts that are linked together by means of specified relationships like 'is a' or 'has'), and (3) knowledge representation using frames (hierarchical networks of collections of attributes in which the topmost elements represent general concepts and the lower elements represent specific instances of the general concepts). Both semantic nets and frames are frame-based.

2000. Among these categories was the development of high-level software, consisting of end user programming, advanced audit software, and expert system development.<sup>4</sup> The Delphi panel predicted that true artificial intelligence with learning capabilities has a 60% probability of being available in the year 2000.

Weber [1978] addressed (among other things) the question of the usefulness of a simulation decision aid in assessing internal control system reliability.<sup>5</sup> In an estimation task regarding the dollar error in inventories, and an audit plan revision task, a group of 40 auditors was split up into one sub-group that used the simulation decision aid and another which did not. The group using the decision aid was more accurate in decision-making, had more confidence in its decisions, took less time, and had a higher level of satisfaction with the available audit evidence than the group that did not use the decision aid.

Mock & Turner [1981] investigated (among other things) the relationship between audit guidance and the sample sizes used in substantive testing. Audit guidance was manipulated through five levels: no guidance, narrative guidance, structured guidance, a statistical approach and manager review. The subjects were 200 auditors from one big-eight U.S. accounting firm. In only eight out of the twenty cases (5 guidance levels times 4 substantive testing procedures) could a significant reduction in sample sizes be observed when the level of guidance increased. Overall guidance effects were significant for three out of the four substantive testing procedures.

Cushing & Loebbecke [1986] compared the audit methodologies of 12 large accounting firms in the U.S.. They hoped to gain a better understanding of the audit process and its practice, which could improve the audit standard-setting process within the AICPA. The audit manuals of the participating accounting firms were studied, a normative audit model was developed, individual firm models were constructed and compared to the normative model, a cross-firm comparison was made based on the compliance of each individual firm model with the normative model, and, finally, the degree of structure within each individual firm approach was assessed. It was found that the participating firms did not completely follow the generally accepted auditing standards on which the normative model was based, and that there were significant differences between firms with regard to the degree of structure.

---

<sup>4</sup> The other categories of technological change were: data communications, information system design, automation of transaction processing, and data security.

<sup>5</sup> The simulation decision aid as used in this experiment was a program, constructed to model the inventory accounting and control system of a manufacturing company. It allowed auditors to change the frequency of errors (and in some cases: the nature of errors) in order to perform what-if analyses. It then simulated a period of transaction processing and produced a list with frequency distributions regarding errors in different inventory accounts.



Kachelmeier & Messier [1990] investigated the effects of the use of a non-statistical decision aid on auditors' sample size decisions. They tested three hypotheses: first, that subjects using the non-statistical decision aid would tend to recommend larger sample sizes than subjects making unaided intuitive sample size decisions, second, that the sample sizes computed by subjects using the decision aid would be systematically lower than the sample sizes implied by the decision aid parameter judgements of a separate group of subjects who were not asked to compute sample sizes, and third, that the use of the decision aid would have no effect on the dispersion of auditors' sample size decisions. They found that the use of the decision aid resulted in systematically larger sample sizes, the sample sizes as calculated from the parameter judgments of the parameters-only group were higher than those computed by the decision aid group. The application of the decision aid was found to have increased the dispersion of sample size responses between auditors.

Wallage [1991] replicated the study of Cushing & Loebbecke [1986]. He investigated the methodologies and the degree of structure of the audit process of 15 accounting firms in the Netherlands.<sup>6</sup> This study investigated the potential methods of analyzing and characterizing audit methodologies and the differences in audit approach between accounting firms with regard to degree of structure. International influences are hypothesized as the most important variables explaining these differences. It appeared that the degree of international influence has a significant effect on the degree to which the audit approach complies with a normative model which was developed. This result implies that the implementation of auditing standards in the Netherlands is based on international standards. Furthermore, it appeared that the partner/qualified auditor ratio explained the degree of structure, whereas the number of offices and the number of qualified auditors had no explanatory power with regard to the degree of structure.

### **Dynamic research: cognitive style**

Driver & Mock [1975] identified five decision style categories. In an experiment in a managerial accounting context,<sup>7</sup> they found significant differences regarding information acquisition, decision quality and decision speed between decision style categories. Weber [1978] investigated the effects of differences in risk aversion, degree of dogmatism and experience. It was found that audit programs become more elaborate as the auditor becomes more risk averse and more experienced. Dogmatism as a personality variable appeared to have no significant effects on decision accuracy, decision confidence, and decision time. Lusk [1979] used an

---

<sup>6</sup> After a number of mergers, 11 firms remained.

<sup>7</sup> Because their study was a pioneering effort, it is discussed here in a review of audit judgment literature.

embedded figures test to identify field-independent and field-dependent subjects.<sup>8</sup> It was found that the high analytics (field-independent) outperformed the low analytics (field-dependent) and that the receivers of elementary reports performed better than the receivers of transformed reports. However, no interaction effect (i.e. differential performance peaking<sup>9</sup>) was observed.

### Static research: lens modelling

Ashton [1974] used six cues as independent variables in a lens model that assessed the strength of internal controls. Decision quality criteria were consistency over time and consensus. No experience effects were observed, but auditors were consistent over time and exhibited consensus. Joyce [1976] found a considerable degree of consensus and self-insight regarding audit program planning decisions made by auditors. Here, consensus was probably due to audit manuals containing detailed prescriptions. Ashton & Kramer [1980] investigated different audit judgment quality measures [Messier, 1983]. Furthermore, they investigated differences between students' and auditors' performance. They found that both groups exhibited consensus, and that both groups had considerable self-insight, but the self-insight of auditors was superior. Ashton & Brown [1980] found a considerable degree of consensus between students and auditors. However, auditors had better self-insight. Mock & Turner [1979, 1981] found a high degree of consensus regarding auditors' reaction to weaker internal controls: sample sizes were increased. Gaumnitz *et al.* [1982] studied the effect of experience and firm affiliation on consensus. They found a high between-auditor consensus, but no experience or firm effects. Hamilton & Wright [1982] investigated the correlation between experience on the one hand and consensus, reliability, and self-insight on the other hand. It was found that experience was negatively correlated with consensus and positively correlated with self-insight. No experience effects on reliability were observed. However, it should be noted that the significance of the results was not very high. This might explain the mixed outcomes. Krogstad *et al.* [1984] investigated what factors determine materiality thresholds. Like Moriarity & Barron [1976, 1979] they found that net income was the main factor that determined the height of materiality thresholds. With regard to experience and consistency/consensus, a positive correlation was found. Frederick & Libby [1986] compared the performance of experienced auditors with the performance of

---

<sup>8</sup> Here again the subjects were not auditors. However, the results were crucial for the further development of cognitive style research aimed at developing tailor-made information systems.

<sup>9</sup> Differential performance peaking in this context means that there is an interaction between the form of an information report and the cognitive style of the subject. High analytics would perform better when provided with transformed reports that must be broken down into elementary pieces (i.e. analyzed). Low analytics would perform better when provided with the elementary information chunks in less-transformed reports. In other words: there are two peaks, one for each group-report combination.

novices. It was found that experienced auditors better understood the relationship between internal control weaknesses and the likelihood of errors. Abdolmohammadi & Wright [1987] found significant differences between experienced and inexperienced subjects in an analysis on three different task levels. However, when pooling all subjects together, a significant experience effect was observed on only one task (out of six). Libby & Frederick [1990] found significant experience effects in the expected direction: experienced subjects gave more plausible explanations for an observed phenomenon, had more accurate perceptions of the relative frequency of occurrence of financial statement errors, would generate more explanations based on highly frequent occurring errors, and generated more explanations from the same transaction cycle as the inherited explanation. Moeckel [1990] investigated the relation between audit experience and the occurrence of memory errors. It was found that experienced auditors more often correctly integrated, more often failed to integrate items on which they had to perform reconstruction processes, and more often followed reconstruction processes.<sup>10</sup> Bedard [1991] studied the relation between domain-specific expertise and quality of audit decision making. It was found that the level of consensus was higher when controls were strong, and lower when controls were weak. Both novices and experts were consistent with professional standards, but only experts were consistent with audit firm standards. Moeckel [1991] found that experienced auditors more often successfully integrated (see also Moeckel [1990]). However, when contradictory evidence came available, the performance of experienced auditors, just like the inexperienced auditors, deteriorated.

### **Static research: probabilistic judgments**

Some earlier probabilistic judgment studies are interesting for the current research since the effects of experience, firm affiliation, and type of education on the assessments of probabilities and degree of consensus between auditors are, incidentally, investigated. Corless [1972] found that the assessment of prior distributions was not affected by the number of years of auditing experience or by educational background. Ward [1976] found a significant, however, not very high consensus regarding materiality assessments between auditors. Lewis [1980] found that consensus, measured as the level of homogeneity of utility functions) was higher in highly material situations than in less material situations. Furthermore he found, that homogeneity was more apparent among members of the same public accounting firm. Felix [1976], like Corless [1972], found that auditors were able to assess prior probabilities fairly well, provided that they had received sufficient training. Crosby [1980] found that auditors were not 'Bayesian'. In a sample size

---

<sup>10</sup> Failure to integrate can be defined as the failure to make mental connections between separately received pieces of information. Reconstruction can be defined as altering the mental representation of information in order to make it consistent with existing knowledge [Moeckel, 1990].

selection task, auditors chose different judgmental sample sizes than Bayes' theorem would predict.

Most studies of probabilistic judgment focus on the applicability of heuristics to auditors' decision-making behavior. Uecker & Kinney [1977] searched for the presence of the representativeness and protectiveness heuristics in auditors' evaluation of random sample results. Only 26% of the auditors did *not* follow any of these heuristics. Implicitly, it can be concluded that there is a great chance that practicing auditors might make judgmental errors, and that care should be taken in applying professional judgment in the evaluation of sample results. Joyce & Biddle [1981a] investigated the use of the anchoring and adjustment heuristic by auditors. Results were mixed: firstly, in fraud estimates anchoring and adjustment was found, secondly, in audit program planning the findings were inconclusive, and thirdly, in a reporting task no anchoring and adjustment was observed. Joyce & Biddle [1981b] tested for the use of the representativeness heuristic among auditors, in order to explain systematic departures from the Bayesian optimum. It was found that representativeness does not completely describe the behavior (consideration of source reliability) of auditors in the experiment. In an extension of the Joyce & Biddle [1981b] study, Bamber [1983] investigated the effects of source reliability on audit judgments. Again, it was found that source reliability is incorporated into the audit judgment process. This finding also, albeit implicitly, supports the statement that the representativeness heuristic does not provide an adequate explanation for auditors' behavior.

The following table gives an overview of audit judgment research.

### Studies of pre-decisional behavior

Weber [1980]	External auditors remembered more EDP controls than internal auditors, who in turn remembered more EDP controls than students.
Biggs & Mock [1980, 1983]	Inexperienced auditors employed a directed information access strategy whereas experienced auditors employed a systemic information access strategy (directed: separate search for information about four interrelated problems; systemic: integral search for information about four interrelated problems in one run).
Biggs, Mock & Watkins [1988]	Both seniors and managers were able to identify the main problems (risks). However, the seniors increased audit work significantly more than managers did.
Johnson, Jamal & Berryman [1989]	The novice auditor did not discover management fraud and unjustly issued an unqualified opinion, whereas the expert auditor did discover the fraud.
Williams [1990]	Risk assessments by auditors are made thoroughly and are adapted to contextual factors. Auditors have developed a cognitive process for making risk estimates which is not audit manual driven.
Mock, Wright, Washington & Krishnamoorthy [1993]	Risk assessment is a very important element in the audit process as observed during experimental problem-solving by experienced auditors.

### Decision aid research

Weber [1978]	Auditors using a simulation decision aid were more accurate in decision-making, had more confidence in their decisions, took less time, and had a higher level of satisfaction with the audit evidence.
Mock & Turner [1981]	Some auditors adjusted sample sizes downward when the level of guidance increased. However, they were outnumbered by the auditors who did not adjust sample sizes when the level of guidance increased.
Cushing & Loebbecke [1986]	There were significant differences between firms with regard to degree of structure.
Kachelmeier & Messier [1990]	Auditors using a decision aid calculated systematically larger sample sizes. Variance increased when using a decision aid.
Murphy [1990]	Nonexpert system users had higher scores in a semantic memory development task and in a combined semantic/episodic memory development task than expert system users (receiving explanations). Nonexpert system users had higher scores on a semantic memory development task than expert system users (not receiving explanations).
Wallage [1991]	International standards influence the degree of structure of audit approaches.

<b>Cognitive style research</b>	
Driver & Mock [1975]	Decisives were the slowest decision-makers, performed poorly, and used few pieces of information. Hierarchics were the fastest decision-makers, performed well, and used many pieces of information. The integratives, complex integrative-hierarchics, and flexible types could be positioned between the decisives and the hierarchics.
Weber [1978]	Audit programs became more elaborate as the auditor became more experienced and risk averse. Decision variability increased with increasing risk-taking propensity. No significant relation between the degree of dogmatism and accuracy, confidence and decision time could be observed.
Lusk [1979]	Differential performance peaking regarding high/low analytics receiving simple/complex reports could not be substantiated.
Benbasat & Dexter [1979]	Differential performance peaking regarding high/low analytics receiving aggregate/disaggregate information was observed.
Pincus [1990]	Field-independent auditors more often judged accounts to be not fairly presented. No significant relation between category width and judgment performance could be observed. High tolerance for ambiguity auditors more often judged accounts to be not materially misstated.
Pincus [1991]	The more experienced auditors were more confident in decisions that would lead to an unqualified opinion and less confident in decisions that would lead to a qualified or adverse opinion. Decision confidence and tolerance for ambiguity were positively related. Auditors with a low risk-taking propensity were more confident in correct decisions, however auditors with a high risk-taking propensity were not more confident in incorrect decisions. Finally, confidence was higher when the decision was in accordance with prior expectations and lower when the decision was not in concordance with prior expectations.
<b>Lens model studies</b>	
Ashton [1974]	Auditors were consistent in their internal control ratings and exhibited consensus. No experience effects were found.
Moriarty & Barron [1976]	The majority of the auditors used additive models to determine materiality as a function of net income, earnings trend and company size. There was no consensus regarding the effect of net income on materiality.
Joyce [1976]	Within firm consensus regarding audit program planning is greater than between firms.
Moriarty & Barron [1979]	There was no consensus regarding materiality thresholds. The strongest effect on materiality came from net income.
Ashton & Kramer [1980]	Consensus and self-insight existed among students as well as among auditors. Auditors exhibited significantly more self-insight. Order of presentation had an equal effect on students' and auditors' rankings of internal controls.

Ashton & Brown [1980]	Students and auditors exhibited consensus and self-insight. Auditors had better self-insight.
Mock & Turner [1979, 1981]	All auditors (consensus) chose higher sample sizes when internal controls were weaker.
Gaumnitz, Nunamaker, Surdick & Thomas [1982]	Auditors exhibited consensus as to the rating of internal controls and planning decisions. They also exhibited a highly consistent behavior as to the ratings of internal controls and the numbers of audit hours required. No differences between firms or experience levels were observed.
Hamilton & Wright [1982]	Experience and consensus were negatively correlated. There was no correlation between experience and judgment stability. There was a positive correlation between self-insight and experience. Differences between novices and experts were not significant.
Krogstad, Ettenson & Shanteau [1984]	Experience was a factor that determined consensus and consistency in judging. Students could not be considered good surrogates for professional auditors.
Libby [1985]	There was a strong relationship between the availability of hypotheses in memory and their perceived frequency of occurrence. Errors which actually are detected more frequently are not judged to occur more frequently. Overstatement errors are judged to occur more frequently than understatement errors. No recency effects were found.
Frederick & Libby [1986]	Expert auditors had a better understanding of relations between internal control weaknesses and the likelihood of error acceptance.
Abdolmohammadi & Wright [1987]	Experienced auditors thought it less necessary to require adjustments or to issue a qualified opinion, calculated smaller sample sizes and considered internal controls to be stronger than did the less experienced auditors.
Libby & Frederick [1990]	Experienced auditors gave more plausible explanations for an observed deviation, had more accurate perceptions of frequencies of error occurrence, generated more errors with higher levels of occurrence, and would generate more explanations from within the same transaction cycle as the inherited explanation.
Moeckel [1990]	Experienced auditors more often successfully integrated information (identification of meaningful relationships between pieces of information), more often failed to integrate information which would have required them to apply reconstructive memory processes, and more often reconstructed memory.
Moeckel [1991]	Experienced auditors more often successfully integrated. However, when contradictory information came available in another audit task or with a different wording, the performance of the more experienced auditors, just like the inexperienced auditors, deteriorated.
Bedard [1991]	Experts exhibited more consensus when internal controls were strong, and less consensus when internal controls were weak. Both novices and experts were consistent with professional standards. Only experts were consistent with accounting firm standards.

Probabilistic judgment models	
Corless [1972]	Differences in the number of years of auditing experience and statistical background did not explain differences in prior probability assessments.
Ward [1976]	Auditors exhibited consensus as to the ranking of items determining materiality.
Felix [1976]	Auditors are able to assess prior probabilities provided that they receive sufficient training. EPS (equivalent prior sampling) as an assessment method looks promising.
Newton [1977]	The more averse the auditor is to accepting beta-risk, the less uncertainty he will accept and the lower his materiality threshold will be.
Uecker & Kinney [1977]	Most auditors (54%) used the representativeness heuristic, whereas a minority (37%) used the protectiveness heuristic. 26% did not make any judgment error, whereas 19% made each error at least once.
Crosby [1980]	Auditors' sample sizes did not correspond with Bayesian sample sizes. Elicitation technique had no influence on outcomes.
Lewis [1980]	The level of homogeneity (i.e. a consensus measure) of utilities is higher in highly material situations than in less material situations. The effect was the strongest among members of the same firms.
Crosby [1981]	Different methods for the elicitation of prior probabilities did not provide significantly inconsistent outcomes: one method could not be considered superior to the other.
Joyce & Biddle [1981a]	In a fraud probability estimation task, the anchor had a significant influence. In internal control quality estimations, the anchor did not have a significant influence, however, an order effect could be observed. In an audit reporting task, no significant effects of experimental manipulation could be observed. However, variance was very high, indicating low consensus.
Joyce & Biddle [1981b]	Auditors' probabilities were regressed towards the base rate. Source reliability was only weighted when its presence was salient.
Bamber [1983]	Auditors considered source reliability an important aspect of their decision-making.
Libby [1985]	There was a strong relationship between the availability of hypotheses in memory and their perceived frequency of occurrence. Overstatement errors are judged to occur more frequently than understatement errors. No recency effects were found.
Mock, Wright, Washington & Krishnamoorthy [1993]	Auditors given compliance test results indicating less reliable data revised their prior probabilities more than auditors given more reliable data. However, in low beta-risk situations, prior probabilities were adjusted more than in high beta-risk situations.

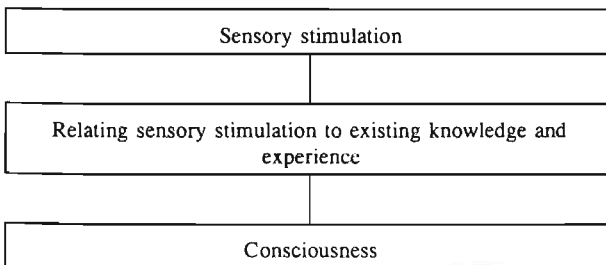


### Historical context of verbal protocol analysis

Think aloud protocols, as derived in VPA, can best be described from a psychological point of view: the thought processes of other people are observed by having them think aloud and by keeping track of physical changes during problem-solving. One of the research methods used in early studies of verbalization was introspection. A brief review of the literature on introspection is therefore necessary before outlining the historical development of VPA, since VPA's 'roots' lie in introspection.

Early research into human information processing was conducted mainly by religiously and philosophically oriented scientists [e.g. James, 1890]. From that background they observed their own cognitive processes, that is applied introspection, as the human mind was generally seen as beyond understanding in scientific terms [Ericsson & Simon, 1984]. Initially introspection was not embedded in a thorough theoretical framework. It was deemed to be as unscientific as the casual observation of natural events would be for the natural sciences. Later a more structured approach to introspection was introduced. Titchener [1912] proposed separating theory from facts by having the subjects describe only their conscious experience, leaving the inferential process to the experimenter. Consciousness was thus seen as being very important when applying introspection and the question of how to report on the contents of consciousness arose. According to Wundt [1893] consciousness (including thought) could be described in terms of sensory components. In figure 1 the relation between the conglomerate of sensations (sensory stimulation) and the concept of consciousness is outlined. For example, in an accounting experiment a sensory stimulation could be: 'seeing that balance sheet total exceeds turnover'. A subject could report on this external stimulation as follows: 'since balance sheet total exceeds turnover the business is probably involved in manufacturing'. This is the sensory component of the stimulus presented.

**Figure 1:** Human experience of external stimuli.



There are several different approaches which can underly the analysis of verbal

reports, and so need to be considered here. Some of the most important of these approaches for the historical development of the study of consciousness have been the structuralist, functionalist, and behaviorist approaches. Structuralism sees consciousness as the base structure of the field investigated by psychology. Structuralism was concerned with the study of human observation, memory structure, thinking structures, etc. Proponents argued that, since there are structural as well as functional elements in biology, and anatomical (structural) as well as physiological (functional) elements in the medical sciences, why then is there not a similar division in psychology? If so, structural psychology should be the basis of functional psychology since the structure must be known before the functioning of the structure can be studied. An analogy can be found in auditing: a thorough description of the administrative organization must be made before compliance testing begins. Proponents of the functionalist view of psychology argue that the functions of the base structure need to be studied rather than the base structure itself. Research based on this view typically studies the adaptation of subjects to their environment by means of their psychological capacities.

Behaviorism emerged as a reaction to structuralism. Functionalism is now regarded as a transitional movement between structuralism and behaviorism, that is between a subjective approach and a more objective approach. The behaviorists rejected introspection as subjective, and instead conducted experiments, as in the natural sciences, in which the observation of subjects' behavior was the main research method. It was therefore not surprising that a behaviorist (J. Watson, in a 1920 article), was one of the first researchers to give a well-documented analysis of thinking aloud activity, and so began the literature on VPA.

Earlier work by Watson [1913], the 'father' of behaviorism, presented the following tenets: first, human psychology as conducted by structuralists is not a natural science. In studying consciousness using introspection, psychology becomes entangled in speculative questions which cannot be experimented upon. Second, psychology, as the behaviorist views it, should be a purely objective branch of the natural sciences, in which introspection may not be used, just as it is not possible to use it in physics or chemistry. Behaviorists assumed that the behavior of animals could be studied without referring to consciousness. Humans and animals exhibit the same behavior, so the study of consciousness is properly a philosophical than a psychological problem. Third, research into the behavior of any animal, even the amoeba, must be considered important from a behaviorist's standpoint. Behavior then, is described in terms of stimulus, response, reinforcement etc. Fourth, the proposed elimination of states of consciousness as research objects would make the gap between psychology and other sciences smaller. Results obtained in psychological experiments could be used as explanations in physiological terms. Fifth, psychology must neglect many phenomena so that the remaining problems can be formulated in such a way that advanced methods, existing or future, will

lead to solutions. Behaviorists thus defined psychology narrowly. For example, if a structuralist wanted to learn more about the auditing concept 'materiality' then he would ask the test person: 'What do you think about the statement that materiality is 1,000,000 ?'. The structuralist would make notes of the answer, thus expanding his knowledge about materiality. On the other hand, the behaviorist would present a business description to a test person together with an amount that could be considered material ( 1,000,000). In his notes the behaviorist would write down exactly how the test person reacted when confronted with the stated materiality in relation to the business description. The structuralist and the behaviorist thus study only a part of the phenomenon; the structuralist could have made notes on the reaction of the test person while the behaviorist could have asked the test person questions.

Here a paradoxical situation emerges: since structuralists and behaviorists could be considered to be opponents, one would not expect to find elements of both approaches in one research method. However this is what we find in VPA. There is no doubt that verbal reports provide introspectively-oriented (structuralist) data, because the test persons involved report on the contents of their thoughts. Behaviorism and allied schools of thought have been schizophrenic about the status of verbalizations as data [Ericsson & Simon, 1980]. The cause of this 'schizophrenia' lies in the perception of what constitutes the core part of the passage: '...test persons report on the contents of their thoughts'. When the part 'contents of thought' is stressed, the researcher typically could be considered to be following a structuralist view. When the part 'test persons report' is stressed, the researcher could typically be considered to be following a behaviorist view.

Initially the term 'verbal reports' referred to the results of introspection as used by structuralists. However a new meaning emerged within the framework of behaviorism. Woodworth [1938] discussed the distinction between expressing thoughts and describing them. Based on this research, a number of studies on the collection of verbal reports were carried out, producing some criticisms of verbal reports as used in behaviorist research [Greenspoon, 1955; Verplanck, 1962; Phelan, 1965; Nisbett & Wilson, 1977]. Among these criticisms were:

- Verbal reports are incomplete, and important behavioral and performance changes are not reflected in them.
- The instruction to subjects to give verbal reports, and the process of giving reports, change the cognitive processes under study.
- The verbal reports are inconsistent with other observable aspects of behavior.
- Verbal reports are unreliable and idiosyncratic, and they do not carry any information which is generalizable or that can further our understanding of performance.

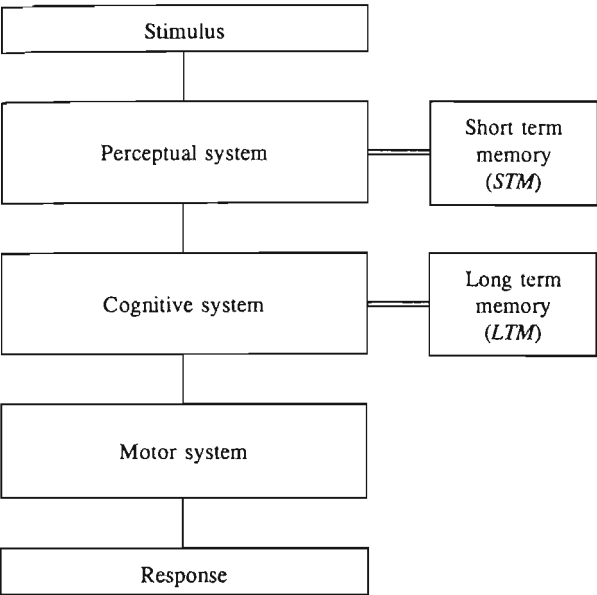
However, many of these criticisms are either unsupported or have been falsified in later research. Flaherty [1974] found that neither overt verbalization nor practice

before real participation significantly influenced problem-solving scores. Hafner [1957] tested a specific application of the verbal hypothesis, which basically states that the acquisition of a motor skill will be facilitated to the extent that verbal description of the stimulus or the response (or both) can be utilized. In a problem-solving situation verbal cues were systematically added without removing other cues. It was found that the addition of verbal cues had only a selective facilitating effect on performance: time consumption was lower and performance was better for the experimental group when compared to the control group. However results were not significant. Ericsson & Simon [1984] support the opinion that there is a preponderance of behaviorist features in VPA, in that it focuses on phenomenological observation rather than on introspection. For example, when the VPA researcher asks his subject to think aloud while or after solving a problem (thus reconstructing his thoughts), only the expressed thoughts, that is behavior, are considered relevant.

Newell, Shaw & Simon [1958] and Newell & Simon [1972] model humans engaged in problem-solving as information processing systems. In this view problem-solving is merely the result of an interaction between information elements in long-term and short-term memory. These memories have different characteristics with regard to capacity and access possibilities. Furthermore the human mind can be considered as a composite of three interacting subsystems: the perceptual system, the motor system, and the cognitive system. The perceptual system consists of sensors which are able to receive information (stimuli), along with related buffer memories (short-term memory: STM) which store the information until it is coded into a format which is readable by the cognitive system. The cognitive system receives coded data from the perceptual system, stores that data in long-term memory (LTM), and retrieves information from long-term memory (experience, existing knowledge) in order to produce output via the motor system, that is to make a decision. The motor system enables execution of a response as a reaction to a stimulus.

Figure 2 gives a graphical representation of the human information processor model. Here the analogy with computerized data processing is evident. The perceptual system can be compared with the man-machine interface of a computer, including the computer's storage of input in internal memory (RAM) in an uncoded format and transmittal to the cognitive system in an encoded format. The cognitive system can be compared with the central processor unit of a computer, as the CPU retrieves and stores data in external memory (disk, tape, eprom) and processes that data to information. The motor system can be compared to the output facility of a computer (screen, printer, or any other device that can be used to output data).

Figure 2: The human information processor model.



According to the human information processor model, human cognition is information processing, that is: a cognitive process can be seen as a sequence of internal states which are subject to changes as a result of information processing. The model can satisfactorily serve as a base for VPA studies because of its high compatibility with various hypotheses which have emerged in the field of human information processing research. High compatibility here implies a certain generalizability, not vagueness in the sense of not being able to capture the results of VPA in computer models of human information processing.

### Myers-Briggs Type Indicator Questionnaire

For each dimension, the Spearman item-total correlation is represented after each item number. Significant ( $p < 0.10$ ) correlations are marked with an asterisk (\*). Respondents were asked to fill out the questionnaire without thinking too long about the answer.

#### Dimension Information Acquisition

1.     0.33     At parties and receptions I am usually:
  - a.     a 'good mixer'.
  - b.     rather quiet and reserved.
2.     0.02     When on an excursion in a plant, I particularly pay attention to:
  - a.     the information that is provided by the guide.
  - b.     the activities of the people who work in the plant.
3.     0.12     Upon arrival in a foreign town, I prefer to:
  - a.     first stay a while in my hotel room.
  - b.     go out and see the town right away.
4.     0.50     \*     I generally get along very well with:
  - a.     people who have imagination.
  - b.     realistic people.
5.     0.11     Personally, I like a novel that:
  - a.     provides a detailed picture of the major character.
  - b.     describes many surprising and unpredictable events.
6.     0.34     When I have been working or studying alone for an hour or so, I:
  - a.     consider a telephone call to be disturbing.
  - b.     would like to talk to somebody.
7.     0.58     \*     I prefer to spend my summer holiday:
  - a.     in a crowded seaside town.
  - b.     in a quiet mountain village.
8.     0.48     \*     When studying, I prefer:
  - a.     a clean desk.
  - b.     everything around.

**Dimension Information Processing**

1. 0.19 When I go somewhere for the day, I prefer:
  - a. to plan what to do and where to go.
  - b. to just go.
2. 0.53 \* Following a schedule:
  - a. appeals to me.
  - b. cramps me.
3. 0.58 \* A planned daily work-schedule:
  - a. provides the best results.
  - b. is not appealing, but necessary.
4. 0.53 \* I am more successful:
  - a. at dealing with the unexpected and seeing quickly what should be done.
  - b. at following a carefully worked out plan.
5. 0.27 When I have a special job to do, I like to:
  - a. organize it carefully before I start.
  - b. find out what is necessary as I go along.
6. 0.34 Problems with personal relations can best be solved in an atmosphere that is:
  - a. rational.
  - b. emotional.
7. 0.08 In the weekend, I prefer to eat:
  - a. whenever I'm hungry.
  - b. at regular times.
8. 0.28 As a teacher, I would emphasize:
  - a. facts.
  - b. theory.

## Appendix F

### Tolerance For Ambiguity Questionnaire

The item-total correlation is represented after each item number. Significant ( $p < 0.01$ ) correlations are marked with an asterisk (\*). Respondents were asked to state whether they agreed or disagreed with each item without thinking too long about the answer.

- |       |                 |   |
|-------|-----------------|---|
| 1.    | 0.30            | A problem has little attraction for me if I don't think it has a solution.  |
| 2.    | 0.35            | I am just a little uncomfortable with people unless I feel that I can understand their behavior.                                  |
| 3.    | 0.34            | There's a right way and a wrong way to do almost everything.  |
| 4.    | 0.34            | I would rather bet on a long shot with high prizes than on a probable winner with only small prizes.                              |
| 5.    | 0.28            | The way to understand complex problems is to be concerned with their larger aspects instead of breaking them into smaller pieces. |
| 6.    | 0.46            | I get pretty anxious when I am in a social situation over which I have no control.  |
| 7.    | 0.34            | Practically every problem has a solution.   |
| 8.    | 0.13            | It bothers me when I am unable to follow another person's train of thought.   |
| 9.    | . <sup>11</sup> | I have always felt that there is a clear difference between right and wrong.  |
| 10.   | 0.23            | It bothers me when I do not know how other people react to me.  |
| 11.   | 0.30            | Nothing gets accomplished in this world unless you stick to some basic rules.   |
| 12.   | 0.28            | If I were a doctor, I would prefer the uncertainties of a psychiatrist to the clear and definite work of a surgeon.               |
| 13.   | 0.16            | Vague and impressionistic pictures really have little appeal for me.  |
| 14.   | 0.24            | If I were a scientist, it would bother me that my work would never be completed because science will always make new discoveries. |
| 15.   | 0.38            | Before an examination, I feel much less anxious if I know how many questions there will be.                                       |
| 16.   | 0.57 *          | The best part of working a jigsaw puzzle is putting in the last piece.  |
| 17. - | 0.02            | Sometimes I rather enjoy going against the rules and doing things I am not supposed to do.  |

---

<sup>11</sup> Since every subject disagreed with this statement, variance is zero and it was not possible to calculate a correlation coefficient.



- 
- |     |      |   |   |
|-----|------|---|---|
| 18. | 0.52 | * | I do not like to work on a problem unless there is a possibility of coming out with a clear-cut and unambiguous answer. |
| 19. | 0.28 |   | I like to fool around with new ideas, even if they turn out later to be a total waste of time.                          |
| 20. | 0.39 |   | Perfect balance is the essence of all good composition.   |

## Appendix G

### Research questions and their interrelationships

#### General research question 1:

Is there a predominant psychological profile among auditors, and is there any relationship between this and number of years audit experience, firm affiliation, and educational background?

Indicators	Psychological profile	Demographics		
		Experience	Educational background	Firm affiliation
MBTI-S	Q1.1a	Q1.2a	Q1.3a	Q1.4a
MBTI-F	Q1.1b	Q1.2b	Q1.3b	Q1.4b
TFAM	Q1.1c	Q1.2c	Q1.3c	Q1.4c

#### General research question 2:

Is there a significant relationship between auditors' psychological profiles and demographic characteristics and their problem-solving behavior?

Demographics and psychological profiles	Problem-solving behavior		
	Operators and task encodings	Recall and completeness of information access	SS/CS/AD
EXP(0,1)	Q2.1a	Q2.2a	Q2.3a
EDU(0,1)	Q2.1b	Q2.2b	Q2.3b
FIRM(1,3)	Q2.1c	Q2.2c	Q2.3c
MBTI-S	Q2.1d	Q2.2d	Q2.3d
MBTI-F	Q2.1e	Q2.2e	Q2.3e
TFAM	Q2.1f1 Q2.1f2	Q2.2f	Q2.3f

#### General research question 3:

What factors determine audit planning stage materiality thresholds?

**VPA Case description Midcom Ltd*****Midcom Ltd***

Midcom Ltd, a producer of advanced communications equipment is involved in a takeover by Comsys Ltd. The parties have agreed that the takeover should take place on 1 January 1989, subject to agreement on the price. The management of Midcom Ltd maintains that the business is worth f4,937,500 as at 1 January 1989 (f=Dutch Guilders). They support this statement with some documents:

1. Balance sheets for December 31<sup>st</sup> of the years 1985-1988.
2. Profit and loss statements for 1985-1988.
3. A number of explanatory notes.

Midcom's annual accounts for 1985-1987 have been audited and received an unqualified opinion. The annual accounts for 1988 have only partially been audited, but Midcom's management have said that they are ready to co-operate with any continuation of the audit by a new auditor.

You have been engaged by Comsys Ltd and instructed to audit Midcom's annual accounts for 1988. You have a limited budget, and are required to carry out your task in the shortest possible time. Every request for information costs extra time, so it is desirable to proceed, as efficiently as possible, on the basis of information placed before you. Midcom's present auditor will be made redundant. His final duties regarded the reconciliation of the 1988 financial statements with the general ledger. Furthermore he checked the existence and completeness of assets and liabilities. The takeover can only proceed if Comsys and Midcom reach agreement on the price, and if Midcom's annual accounts receive an unqualified opinion.

Note: 'audit value', in this case description, means the value according to the auditor.

**Problem:**

1. How do you evaluate the materiality of the various items in the balance sheet and the profit and loss account, in other words, what factors are relevant in passing judgment on the reliability of the annual accounts?
2. What quantitative standards of materiality would you use in this audit?
3. What additional information would you like to have had access to, when certifying the annual accounts?

These questions do not have to be approached in this order. In practice auditing is an iterative process, in which one can reconsider decisions taken earlier. Type [return] for the main menu (i.e. the information page which shows the main classification of the data). From the main menu you can access various sub-menus or information pages. Below each page of information the options for further choices are listed. All amounts are in thousands of guilders, except for those in the 'audit value' summary, and the summary of investments.

### Main Menu

Make a choice by typing the number which appears before the further information desired, or type [return].

- 1           Balance sheets
- 2           Profit and loss accounts
- 3           Supplementary notes
- 4           Investments
- 5           Management
- 6           Contracts
- 7           Internal controls and audit evidence
- 8           Exit
- [return]   Return to first introductory page

## Appendix I

### Information page standard ranking, average ranking, and attention duration

The standard ranking is based on the top-down access whereas the average ranking is based on the average ranking of items over auditors<sup>12</sup>.

Standard ranking	Average ranking	Page	Average attention duration in seconds	Brief description
1	18	P1	27.63	Assets 1985
2	25	P10	10.33	Liabilities 1985
3	19	P2	21.42	Assets 1986
4	21	P11	14.54	Liabilities 1986
5	24	P3	12.29	Assets 1987
6	26	P12	6.96	Liabilities 1987
7	1	P4	199.08	Assets 1988
8	3	P13	172.63	Liabilities 1988
9	23	P5	15.38	P/L 1985
10	8	P6	29.50	P/L 1986
11	17	P8	33.67	P/L 1987
12	2	P9	193.42	P/L 1988
13	4	P17	108.62	Valuation
14	12	P18	50.17	Personnel
15	9.5	P19	164.62	Market expectations
16	11	P20	112.13	Justification selling price
17	5	P14	69.46	Investments
18	6	P15	78.96	Management
19	7	P16	133.04	Contracts
20	13	P21	64.03	Credit limits
21	9.5	P22	114.21	Setup internal controls
22	14	P23	158.00	Results compliance tests
23	16	P24	103.17	Results substantive tests
24	15	P25	131.00	Audit values ass. 31/12/88
25	22	P29	138.96	Audit values liab. 31/12/88
26	20	P26	152.29	Audit values P/L 1988

<sup>12</sup> Since subject 5 was not incorporated into the analyses, the average ranking and attention duration has been determined on the basis of the ranks assigned by and the attention durations of the remaining 24 subjects.

## Appendix J

### Protocol encoding

One subject verbalized the following during a thinking aloud session:

'I look at the balance sheet in order to get any idea about what we are talking about. How big is this business and what are the balance sheet proportions? What is the main activity of this company? I look at the financial structure. I look at the profit and loss accounts of preceding years in order to assess whether the company is rapidly growing. I catch myself looking at those figures from the standpoint of a buying party advisor. If it is a rapidly growing business then where must I focus in my advisory task, what advice must be given with caution? Those are all matters that have nothing to do with materiality. I get distracted by the very broad role which one has being an advisor in a takeover situation, whereas this case is about materiality. I see a profit and loss account with a turnover. According to the rules that we know, or rules that are given by a global accounting firm guideline, a turnover of 34 million, assuming that the unit of measurement is millions, that's 250.000,- guilders plus a small percentage of these four million, which you can neglect. I would start with a planning stage materiality of 250,000 based on the volume of this business.'

Encoder 1 found the following relevant statements in this transcription:

1. The balance sheet is relevant for general insight into the magnitude and the nature of the business, balance sheet proportions and the financial structure.
2. Profit and loss accounts of former years give an insight into the growth rate of the business.
3. The financial data are mainly looked upon from the standpoint of a buying party advisor.
4. The financial data of the company are of only limited importance as to materiality.
5. Based on an internal guideline, this company having a turnover of 34 million, planning stage materiality is 250,000.

Encoder 2 found the following relevant statements in this transcription:

1. The balance sheet is relevant for obtaining insight into balance sheet proportions, business activities, and financial structure.
2. The profit and loss accounts of preceding years indicate the growth rate of the business.
3. The broad role of the advisor in a takeover, and the aspects that he focuses on, have nothing to do with materiality.
4. According to an accounting firm guideline, a turnover of 34 million leads to a materiality threshold of 250,000.

Based on the results as represented in the table, the percentage of identical perceptions regarding relevance of verbalizations is calculated as follows:

$$percentage = agree / total\ number\ of\ statements$$

If both encoders agree on a relevant statement (code=1), then the variable *agree* is increased by one.

Hence, in this sample the percentage of identical perceptions regarding relevance of verbalizations is 50% (statements 1, 2, and 5 were considered relevant by both encoders).

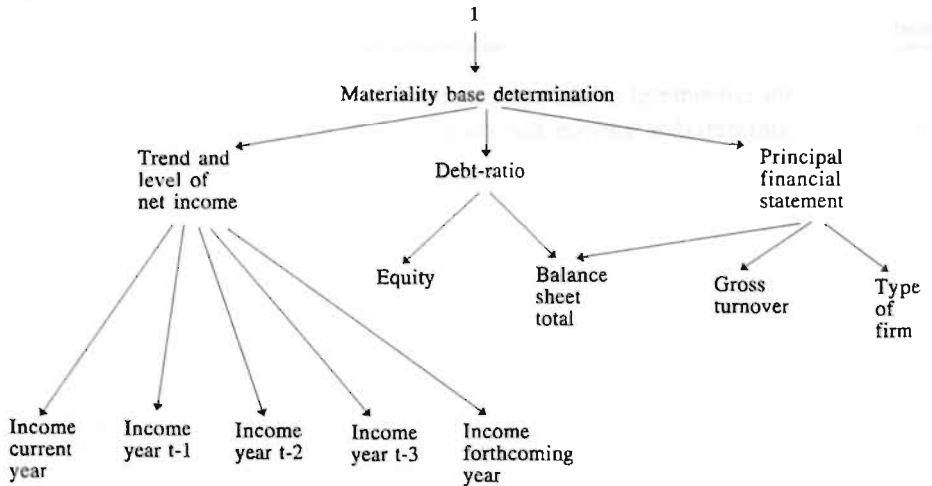
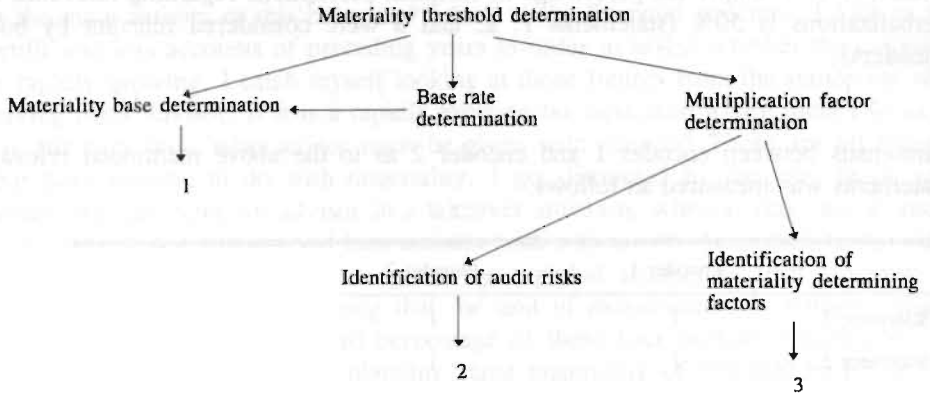
Consensus between encoder 1 and encoder 2 as to the above mentioned relevant statements was measured as follows:

	Encoder 1	Encoder 2	Agree
Statement 1	1	1	1
Statement 2	1	1	1
Statement 3	1	0	0
Statement 4	1	0	0
Statement 5	1	1	1
Statement 6	0	1	0

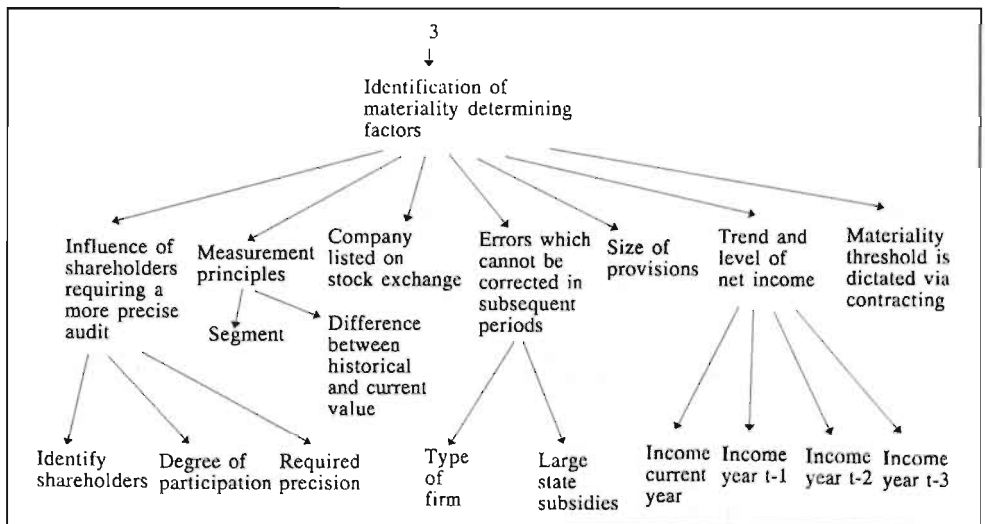
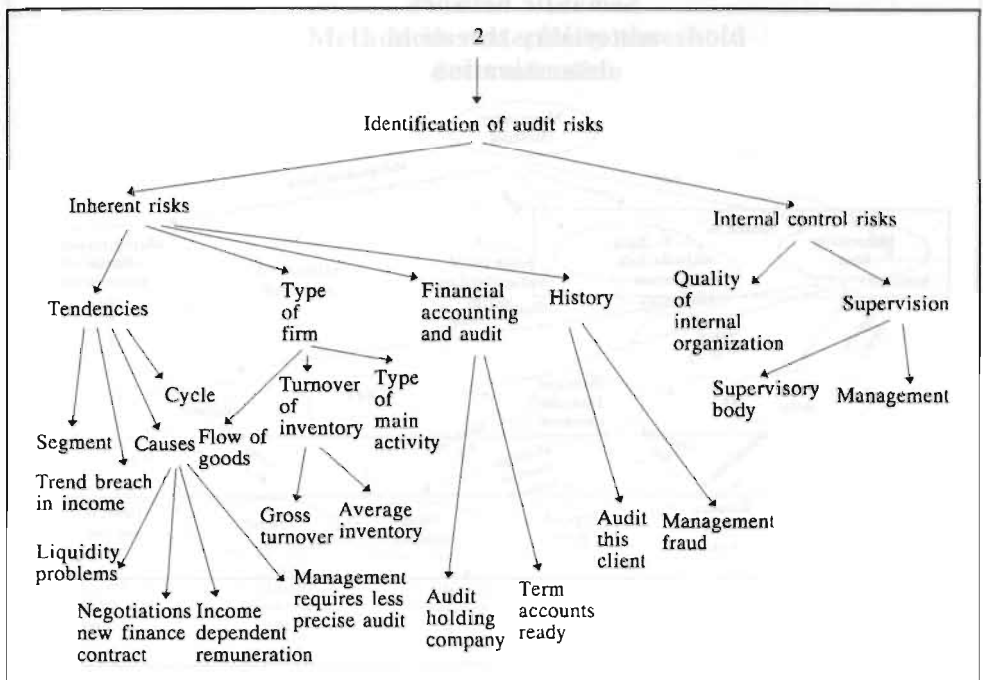
- 0 = not considered a relevant statement.
- 1 = considered a relevant statement.

## Appendix K1

### Task structure materiality threshold determination

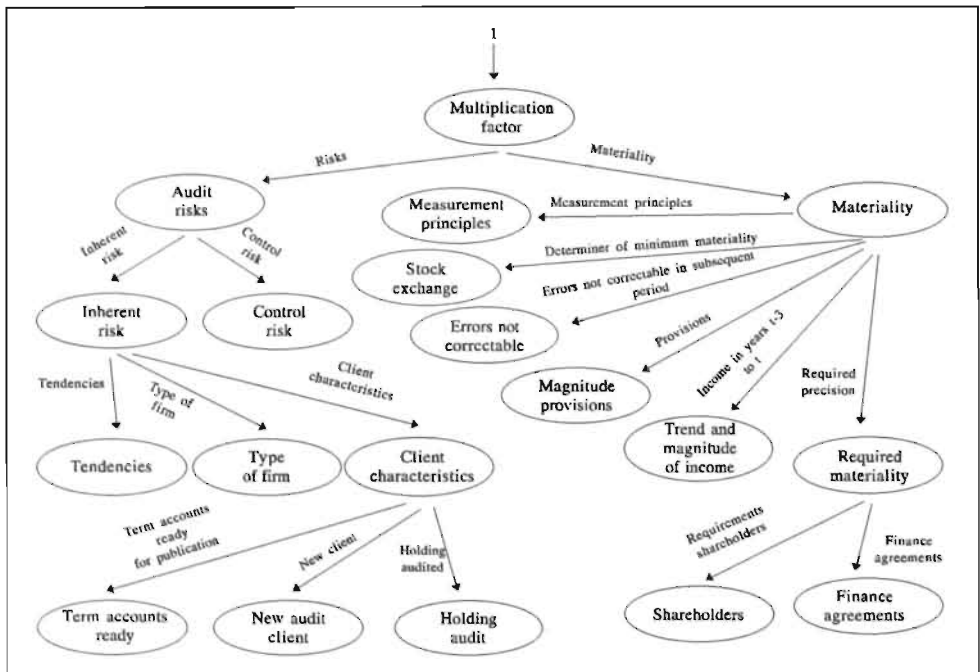
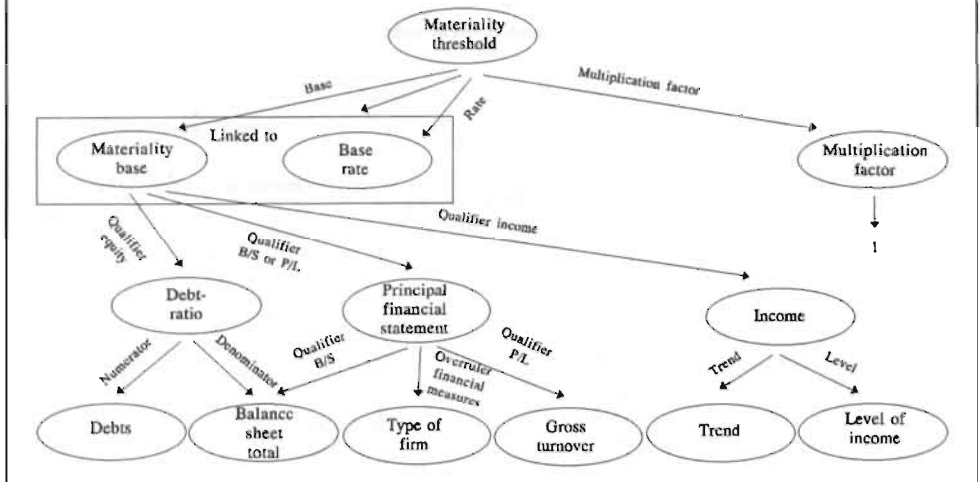






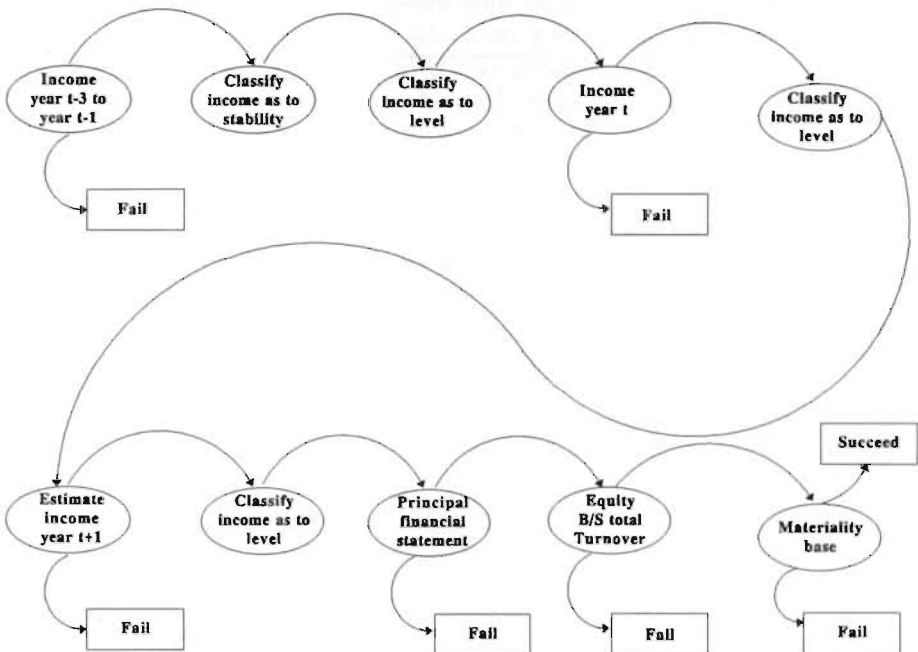
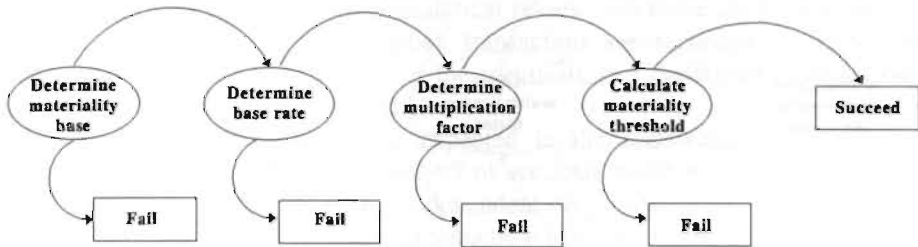
## Appendix K2

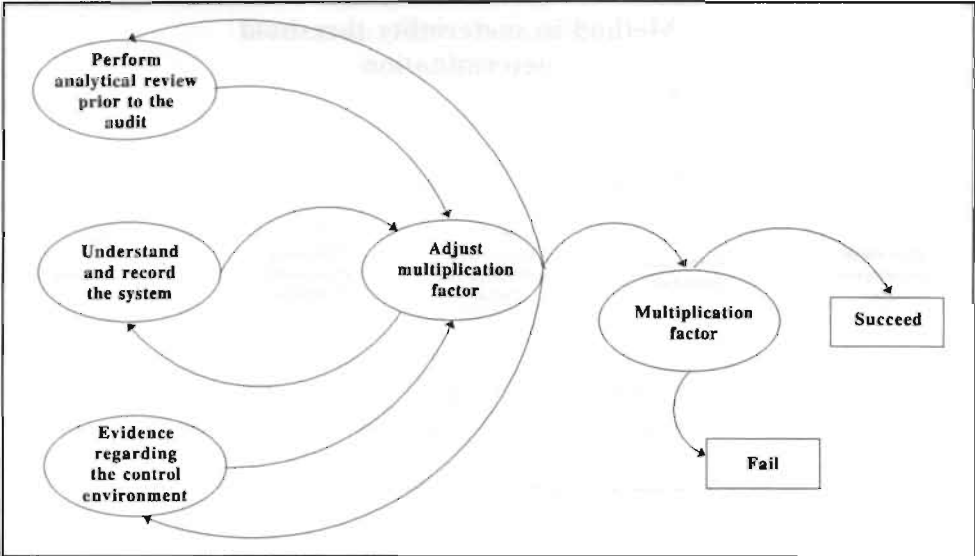
### Semantic network materiality threshold determination



### Appendix K3

#### Method in materiality threshold determination





## Structured cases used for system examination

### *Production Ltd*

Your public accounting firm has been auditing the financial statements of Production Ltd (computer equipment) for 23 years. After reviewing previous years' audit files and conducting an initial analytical review you make the following notes: There is some evidence that fictitious transaction are recorded (in non-routine cycles): according to one employee of the administrative department, some general ledger entries could not be validated.

Other misrepresentations are to be expected in the acquisition and expenditure cycle, with regard to accuracy and cut-off of accounts payable.

Management remuneration is partially dependent on profits.

There is an organizational climate that somehow tolerates misrepresentations.

There is an outside board of directors.

The company is not listed on the stock exchange.

Shareholders are dominant in discussions with outside directors.

There are 6 main shareholders, 3 of which are part of the management, 2 of which are relatives of the management and try to influence managerial decision-making more and more, 1 is an international holding company which owns less than 30%.

This holding company does not have (and does not want to have) a substantial influence on policy.

The international company is audited by the same public accounting firm.

The bank with whom negotiations about a new financing contract (based on reported equity and profitability) have just started requires the financial statements to be more precisely determined than in the preceding years, however management is inclined to negotiate about the audit fee and hence demands a less precise audit.

In the past, no problems regarding financing contract requirements could be observed.

The (trustworthy) management has been leading the company for 23 years.

Net income has been stable at a relatively high level for the last 5 years.

Financial statements are mainly based on historical costs, however some accounts are based on current prices. The difference between historical costs and current prices has been very small the last 5 years.

*Balance sheet 31/12 1990:***Assets**

Fixed assets	9,269,500
Current assets	7,998,000
Liquidities	<u>2,300,000</u>
	19,567,500

**Liabilities**

Equity	10,500,000
Long term debts	7,567,500
Short term debts	<u>1,500,000</u>
	19,567,500

*Profit and loss account 1990:*

Gross turnover	23,785,000
Inventory change	<u>500,000</u>
	23,285,000
Costs	<u>14,790,150</u>
Income before tax	8,494,850
Tax	<u>3,500,000</u>
Net income	4,994,850

Fixed assets consist of the production and office buildings. Current assets consist of accounts receivable (less 15% provision for bad debt) and inventories (less 1% provision for obsolescence).

Any error not corrected in the current year must be corrected in the subsequent year or the year of detection.

Analytical review does not reveal remarkable fluctuations in any of the reported figures.

The administration functions well and prepares the financial statements within 1 month after end of bookyear.

***Wholesale Ltd***

Your public accounting firm has been auditing the financial statements of Wholesale Ltd (kitchen utensils) for 23 years. After reviewing prior years' audit files and conducting an initial analytical review you make the following notes:

There is no evidence that fictitious transactions are recorded or any (material) misrepresentations are to be expected.

The internal organization is acceptable.

Management is receiving a remuneration that is partially dependent on profits.

There is no supervisory body, such as an outside board of directors.

The company is not negotiating a new financing contract.

There are no participants who demand a more or a less precise audit.

The company is not listed on the stock exchange.

Management has a good reputation.

Net income has been growing the last 3 years and it is expected that it will reach a relatively high level in the forthcoming year.

The shareholders are not really involved with the company; only profitability counts (none of the shareholders is part of the management).

The shares are not (partially) owned by a holding company.

Financial statements are based on historical costs.

*Balance sheet 31/12 1990:***Assets**

Fixed assets	2,123,500
Current assets	8,999,000
Liquidities	<u>3,456,000</u>
	14,578,500

**Liabilities**

Equity	6,500,000
Long term debts	7,567,500
Short term debts	<u>511,000</u>
	14,578,500

*Profit and loss account 1990:*

Gross turnover	23,785,000
Costs of sales	<u>19,923,000</u>
	3,862,000
Other costs	<u>1,125,500</u>
Income before tax	2,736,500
Tax	<u>1,279,500</u>
Net income	1,457,000

Fixed assets consist of the warehouse and office buildings. Current assets consist of accounts receivable (less 15% provision for bad debt) and inventories (less 15% provision for obsolescence).

If the assets were based on current prices, then a completely different balance sheet would be presented.

There are no potential errors that cannot be corrected in subsequent years.

Analytical review did reveal some remarkable fluctuations.

The financial statements were traditionally ready just in time.



# Bibliography

Abdel-khalik, A.R., Solomon, I., (eds.), *Research Opportunities in Auditing: the second Decade*, American Accounting Association: Auditing Section, Sarasota, 1988

Abdolmohammadi, M., Wright, A., An Examination of the Effects of Experience and Task Complexity on Audit Judgments, *The Accounting Review*, Vol. 62, No. 1, January, 1987

AICPA, *Professional Standards*, U.S. Auditing Standards, Volumes 1/2, Commerce Clearing House, Inc., June, 1988

Akresh, A.D., Observations on Needed Research in Auditing: Materiality and Audit Risk, *Advances in Accounting*, Vol. 1, 1984

Antle, R., The Auditor as an Economic Agent. *Journal of Accounting Research*, Vol. 20, No. 2, Pt. II, Autumn, 1982

Anzai, Y., Doing, Understanding, and Learning in Problem Solving, in: *Production System Models of Learning and Development*, Klahr, Langley, and Neches, MIT Press, Cambridge, 1987

American Accounting Association, 'Report of the Committee on Basic Auditing Concepts', *The Accounting Review*, Supplement, 1972

Anderson, M.J., Some Evidence on the Effect of Verbalization on Process: A Methodological Note, *Journal of Accounting Research*, Vol. 23, No. 2, Autumn, 1985

Arens, A.A., Loebbecke, J.K., *Auditing, An Integrated Approach*, Prentice-Hall, Englewood Cliffs, 5th ed., 1991

Ashton, A.H., Does Consensus Imply Accuracy in Accounting Studies of Decision Making?, *The Accounting Review*, Vol. 60, 1985

Ashton, A.H., Experience and Error Frequency Knowledge as Potential Determinants of Audit Expertise, *The Accounting Review*, Vol. 66, No. 2, April, 1991

Ashton, R.H., An Experimental Study of Internal Control Judgments, *Journal of Accounting Research*, Spring, 1974

Ashton, R.H., *Human Information Processing in Accounting*, Studies in Accounting Research No. 17, American Accounting Association, 1982

Ashton, R.H., *Some Early Contributions to the Study of Audit Judgment*, Garland, New York, 1984

Ashton, R.H., Brown, P.R., Descriptive Modeling of Auditors' Internal Control Judgments: Replication and Extension, *Journal of Accounting Research*, Spring, 1980

Ashton, R.H., Kramer, S.S., Students as Surrogates in Behavioral Research: Some Evidence, *Journal of Accounting Research*, Spring, 1980

Bamber, M.E., *Expert Judgment in the Audit Team: An Examination of Source Credibility*, Ph.D. Dissertation, Ohio State University, 1980

Bamber, M.E., Expert Judgment in the Audit Team: A Source Reliability Approach, *Journal of Accounting Research*, 21, 1983

- Bamber, E., Snowball, D., An Experimental Study of the Effects of Audit Structure in Uncertain Task Environments, *The Accounting Review*, 63, July, 1988
- Bamber, E., Tubbs, R., Audit Structure and its Relation to Role Conflict and Role Ambiguity: an Empirical Investigation, *The Accounting Review*, 64, April, 1989
- Bamber, E., Bamber, L., Schoderbeck, M., Audit Structure and Other Determinants of Audit Report Lag: an Empirical Analysis, *Auditing: a Journal of Practice and Theory*, 12, Spring, 1993
- Bar-Hillel, M., The Role of Sample Size in Sample Evaluation, *Organizational Behavior and Human Performance*, October,
- Bar-Hillel, M., *Representativeness Reconsidered*, Decision Research Working Paper, 1980
- Batenburg, P.C. van, Kriens, J., Bayesian Discovery Sampling: a Simple Model of Bayesian Inference in Auditing, *The Statistician*, 38, 1989
- Bedard, J., Expertise in Auditing: Myth or Reality?, *Accounting, Organizations and Society*, Vol. 14, Nos. 1/2, 1989
- Bedard, J., *Source Reliability and Audit Evidence Evaluation: A Decision under Ambiguity Analysis*, working paper, Université Laval, 1990
- Bedard, J., *Unreliable Evidence in Auditing: The Problem, its Impact and Solutions*, paper presented at the 14th Annual Congress of the European Accounting Association, Maastricht, The Netherlands, 1991a
- Bedard, J., Expertise and its Relation to Audit Decision Quality, *Contemporary Accounting Research*, Vol. 8, No. 1, Fall, 1991b
- Bedard, J.C., A Multiple Criteria Model for Audit Planning decisions, *Contemporary Accounting Research*, Vol. 8, No. 1, Fall, 1991
- Bedard, J.C., Biggs, S.F., Pattern Recognition, Hypothesis Generation, and Auditor Performance in an Analytical Task, *The Accounting Review*, Vol. 66, No. 3, July, 1991
- Bedard, J., Chi, M.T.H., 'Expertise', *Working paper*, Learning Research and Development Center, University of Pittsburgh, 1991
- Benbasat, I., Dexter, A., Value and Events Approaches to Accounting: An Experimental Evaluation, *The Accounting Review*, October, 1979
- Bernstein, L.A., The Concept of Materiality, *The Accounting Review*, 42, no.1, January, 1967
- Bhaskar, R., Simon, H.A., Problem Solving in Semantically Rich Domains: an Example from Engineering Thermodynamics, *Cognitive Science*, 1, 1977
- Biggs, S., Buijink, W., Maijoor, S., Mock, T., Quadackers, L., Schilder, A., An Assessment of the Relevance and Respectability of Audit Research: Reflections on Reflections of Bindenga, *De Accountant*, februari, 1994
- Biggs, S.F., Selfridge, M., *GC-X: a Prototype Expert System for the Auditor's Going Concern Judgment*, University of Southern California Audit Judgment Symposium, Los Angeles, February, 1986

- Biggs, S.F., Mock, T.J., *Criteria Auditors Use in the Evaluation of Internal Accounting Controls*, Working Paper, Center for Accounting Research, University of Southern California, August, 1979
- Biggs, S.F., Mock, T.J., *Auditor Information Search Processes in the Evaluation of Internal Controls*, working paper, University of Wisconsin-Madison, February, 1980
- Biggs, S.F., Mock, T.J., *Detailed Investigation of Auditor Decision Processes in the Evaluation of Internal Controls and Audit Scope Decisions*, working paper, Center for Accounting Research, University of Southern California, July, 1981
- Biggs, S.F., Mock, T.J., An Investigation of Auditor Decision Processes in the Evaluation of Internal Controls and Audit Scope Decisions, *Journal of Accounting Research*, Spring, 1983
- Biggs, S.F., Mock, T.J., Watkins, P.R., Auditors' Use of Analytical Review in Audit Program Design, *The Accounting Review*, Vol. LXIII, No.1, January, 1988
- Bindenga, A.J., Research voor Accountants, Overpeinzingen na een Symposium, *De Accountant*, september, 1993
- Blaylock, B.K., Rees, L.P., Cognitive Style and the Usefulness of Information, *Decision Sciences*, Vol. 15, 1984
- Bonner, S.E., Experience Effects in Auditing: the Role of Task-Specific Knowledge, *The Accounting Review*, Vol. 65, No. 1, January, 1990
- Bonner, S., Is Experience Necessary in Cue Measurement? The Case of Auditing Tasks, *Contemporary Accounting Research*, Vol. 8, No. 1, Fall, 1991
- Bonner, S.E., Lewis, B.L., Determinants of Auditor Expertise, *Journal of Accounting Research*, Vol. 28, Supplement, 1990
- Bonner, S.E., Pennington, N., Cognitive Processes and Knowledge as Determinants of Auditor Expertise, *Journal of Accounting Literature*, Vol. 10, 1991
- Bonner, S.E., Walker, P.L., The Effects of Instruction and Experience on the Acquisition of Auditing Knowledge, *The Accounting Review*, Vol. 69, No. 1, January, 1994
- Boritz, J.E., The Effect of Research on Audit Planning and Review Judgments, *Journal of Accounting Research*, Autumn, 1986
- Boshuizen, H.P.A., *De Ontwikkeling van Medische Expertise, Een Cognitief-Psychologische Benadering*, dissertation, Rijksuniversiteit Limburg, Maastricht, 1989
- Breuker, J.A., Elshout, J.J., Someren, M.W. van, Wielinga, B.J., Hardopdenken en Protokolanalyse, *Tijdschrift voor Onderwijsresearch*, 11, nr. 5, 1986
- Brown, R.G., Objective Internal Control Evaluation, *Journal of Accountancy*, November 1962
- Brown, S., *Assessing Professional Expertise - DBMS versus Expert System: a Comparative Study*, The Institute of Chartered Accountants of Scotland, 1987
- Brunswik, E., *The Conceptual Framework of Psychology*, University of Chicago Press, Chicago, 1952

- Butt, J.L., Frequency Judgments in an Auditing-Related Task, *Journal of Accounting Research*, Vol. 26, No. 2, Autumn, 1988
- Casey, C.J., Variation in Accounting Information Load: The Effect on Loan Officers' Predictions of Bankruptcy, *The Accounting Review*, January, 1980
- Chandler, J.S., *Expert systems in Auditing: Lessons from the Eighties for the Nineties*, paper presented to: 'Using Expert Systems by Accountants and Auditors'; an International Conference Organized by The Committee on Auditing and EDP of NIVRA, Amsterdam, The Netherlands, September, 1988
- Chenhall, R., Morris, D., The Effect of Cognitive Style and Sponsorship Bias on the Treatment of Opportunity Costs in Resource Allocation Decisions, *Accounting, Organizations and Society*, Vol. 16, No. 1, 1991
- Choo, F., *Expert-Novice Differences in Judgment/Decision Making research*, *Journal of Accounting Literature*, Vol. 8, 1989
- Choo, F., Trotman, K.T., The Relationship Between Knowledge Structure and Judgments for Experienced and Inexperienced Auditors, *The Accounting Review*, Vol. 66, No. 3, July, 1991
- Chow, C., Kramer, L., Wallace, W., The Environment of Auditing, in: Abdel-Khalik, A.R., Solomon, I. (eds.), *Research Opportunities in Auditing: the second Decade*, AAA, Sarasota, pp.155-183, 1988
- Christ, M.Y., Evidence on the Nature of Audit Planning Problem Representations: an Examination of Auditor Free Recalls, *The Accounting Review*, Vol. 68, No. 2, April, 1993
- Church, B.K., Auditors' Use of Confirmatory Processes, *Journal of Accounting Literature*, Vol. 9, 1990
- Colbert, J.L., The Effect of Experience on Auditors' Judgments, *Journal of Accounting Literature*, Vol. 8, 1989
- Connell, N.A.D., Expert Systems in Accountancy: A Review of Some Recent Applications, *Accounting and Business Research*, Vol. 17, No. 67, 1987
- Corless, J., Assessing Prior Distributions for Applying Bayesian Statistics in Auditing, *The Accounting Review*, July, 1972
- Crosby, M.A., Implications of Prior Probability Elicitation on Auditor Sample Size Decisions, *Journal of Accounting Research*, Vol. 18, No. 2, Autumn, 1980
- Crosby, M.A., Bayesian Statistics in Auditing: A Comparison of Probability Elicitation Techniques, *The Accounting Review*, Vol. LVI, No. 1, April, 1981
- Cushing, B.E., Searfoss, D.G., Randall, R.H., Materiality Allocation in Audit Planning: A Feasibility Study, *Journal of Accounting Research*, Vol. 17, Supplement, 1979
- Cushing, B.E., Searfoss, D.G., Randall, R.H., A Reply to Discussions (Loebbecke & Ward, 1979; Scott, 1979) of Materiality Allocation in Audit Planning: A Feasibility Study (Cushing & Searfoss, 1979), *Journal of Accounting Research*, Vol. 17, Supplement, 1979
- Cushing, B.E., Loebbecke, J.K., *Comparison of Audit Methodologies of Large Accounting Firms*, American Accounting Association, Studies in Accounting Research, No. 26, 1986

- Davis, D.L., Grove, S.J., Knowles, P.A., An Experimental Application of Personality Type as an Analogue for Decision-making Style, *Psychological Reports*, 66, 1990
- Davis, J.S., Solomon, I., *Experience, Expertise, and Expert-Performance Research in Public Accounting*, Journal of Accounting Literature, Vol. 8, 1989
- Demski, J.S., Swieringa, R.J., Discussion of Behavioral Decision Theory: Processes of Judgment and Choice, *Journal of Accounting Research*, Vol. 19, No. 1, Spring, 1981
- Dijk, J.C. van, *Expertsystemen voor Accountants*, Delwel, Den Haag, 1990
- Dillard, J.F., Mutchler, J.F., *Knowledge Based Expert Computer Systems for Audit Opinion Decisions*, University of Southern California Audit Judgment Symposium, Los Angeles, February, 1986
- Doktor, R.H., Hamilton, W.F., Cognitive Style and the Acceptance of Management Science Recommendations, *Management Science*, April, 1973
- Driver, M.J., Mock, T.J., Human Information Processing, Decision Style Theory and Accounting Information Systems, *The Accounting Review*, July, 1975
- Dungan, C.W., *A Model of an Audit Judgment in the Form of an Expert System*, dissertation, University of Illinois, UMI, 1983
- Ebert, R.J., Kruse, T.E., Bootstrapping the Security Analyst, *Journal of Applied Psychology*, February, 1978
- Einhorn, H.J., Expert Judgment: Some Necessary Conditions and an Example, *Journal of Applied Psychology*, 59, 1974
- Einhorn, H.J., Hogarth, R.M., Behavioral Decision Theory: Processes of Judgment and Choice, *Journal of Accounting Research*, Vol. 19, No. 1, Spring, 1981
- Ericsson, K.A., Simon, H.A., *Protocol Analysis: Verbal Reports as Data*, The MIT Press, Cambridge, Massachusetts, 1984
- Ericsson, K.A., Simon, H.A., Verbal Reports as Data, *Psychological Review*, 87, 3, 1980
- Feenstra, D.W., *De Beste Stuurder*, inaugural lecture, delivered in a shorter version on taking up the chair of professor in business economics at the Open University, the Netherlands Wolters-Noordhoff, Groningen, 1988
- Feenstra, D.W., Poel, J.H.R. van de, *Bedrijfseconomie, Methodologie en Besluitvorming*, deel I, II en III, *Maandblad voor Accountancy en Bedrijfshuishoudkunde*, 1985
- Felix, W.L., Evidence on Alternative Means for Assessing Prior Probability Distributions for Audit Decision Making, *The Accounting Review*, Vol. LI, No. 4, October, 1976
- Felix, W.L., Kinney, W.R. Jr., Research in the Auditor's Opinion Formulation Process: State of the Art, *The Accounting Review*, vol. 57, No. 2, April 1982
- Financial Accounting Standards Board, *Statement of Financial Accounting Concepts No. 2, Qualitative Characteristics of Accounting Information*, FASB, May 1980

Flaherty, E.G., The Thinking Aloud Technique and Problem Solving Ability, *Journal of Educational Research*, Vol. 68, 1974

Fong Chua, W., Laughlin, R.C., Lowe, E.A., Puxty, A.G., *Four Perspectives on Accounting methodology*, paper presented at the workshop on Accounting and Methodology, European Institute for Advanced Studies in Management, Brussels, December, 1981

Frederick, D.M., Auditors' Representation and Retrieval of Internal Control Knowledge, *The Accounting Review*, Vol. 66, No. 2, April, 1991

Frederick, D., Libby, R., Expertise and Auditors' Judgments on Conjunctive Events, *Journal of Accounting Research*, Autumn, 1986

Gal, G., *Using Auditor Knowledge to Formulate Data Model Constraints: An Expert System for Internal Control Evaluation*, Ph.D dissertation, Michigan State University, 1985

Gaumnitz, B.R., Nunamaker, T.R., Surdick, J.J., Thomas, M.F., Auditor Consensus in Internal Control Evaluation and Audit Program Planning, *Journal of Accounting Research*, Vol. 20, No. 2, Autumn, 1982

Gibbins, M., *Human Inference, Heuristics and Auditors' Judgment Processes*, paper presented at the Canadian Institute of Chartered Accountants Auditing Research Symposium, Laval University, November, 1977

Gibbins, M., Propositions about the Psychology of Professional Judgment in Public Accounting, *Journal of Accounting Research*, Vol. 22, No. 1, Spring 1984

Glazer, R., Steckel, J.H., Winer, R.S., Locally Rational Decision Making: the Distracting Effect of Information on Managerial Performance, *Management Science*, Vol. 38, No. 2, February, 1992

Glover, J.A., Ronning, R.R., Bruning, R.H., *Cognitive Psychology for Teachers*, MacMillan Publishing Company, New York, 1990

Gorry, G., Scott-Morton, M., A Framework for Management Information Systems, *Sloan Management Review*, Fall, 1971

Graham, L.E., Damens, J., Ness, G. van, *Developing Risk Advisor: An Expert System for Risk Identification*, working paper, Coopers & Lybrand, 1990

Greenspoon, J., The Reinforcing Effect of Two Spoken Sounds on the Frequency of Two Responses, *American Journal of Psychology*, 68, 1955

Grötschel, M., Wakabayashi, Y., *A Cutting Plane Algorithm for a Clustering Problem*, Institut für Mathematik, Universität Augsburg, Schwerpunktprogramm der Deutschen Forschungsgemeinschaft 'Anwendungsbezogene Optimierung und Steuerung', Report No. 9, 1987

Grudnitsky, G., *A Prototype of an Internal Control Expert System for the Sales/Accounts Receivable Application*, University of Southern California Audit Judgment Symposium, Los Angeles, February, 1986

Gul, F.A., The Joint and Moderating Role of Personality and Cognitive Style on Decisionmaking, *The Accounting Review*, No. 2, 1984

Gwilliam, D., *A Survey of Auditing Research*, Research Studies in Accounting, The Institute of Chartered Accountants in England and Wales, Prentice Hall International, 1987

- Hafner, A.J., Influence of Verbalization on Problem Solving, *Psychological Reports*, 3, 1957
- Hamilton, R.E., Wright, W.F., Internal Control Judgments and Effects of Experience: Replications and Extensions, *Journal of Accounting Research*, Vol. 20, No. 2, Autumn, 1982
- Hanssen, J.V., Messier, W.F., *A Preliminary Test of EDP-XPert*, University of Southern California Audit Judgment Symposium, Los Angeles, February, 1986
- Hayes, R.S., *Proposal for a Computerized Audit Support System*, Rijksuniversiteit Limburg, April 1988
- Hayes, R.S., *The Utilisation of Accounting Information by International Bank Loan Officers in Their Loan Decisions: A Verbal-Protocol-Based Expert System Simulation*, Ph.D. Dissertation, London School of Economics and Political Science, Departments of Accounting & Finance and Information Systems, 1989
- Hayes, R.S., Poel, J. van de, *Methodologies of Expert Modelling: A Critical Review of 21 Accounting and Auditing Expert Systems*, paper presented at the 13th Annual Congress of the European Accounting Association, Budapest, Hungary, 1990
- Heimann, S.R., Chesley, G.R., Audit Sample Sizes for Aggregated Statement Accounts, *Journal of Accounting Research*, Autumn, 1977
- Henderson, J.C., Nutt, P.C., The Influence of Decision Style on Decision Making Behavior, *Management Science*, Vol. 26, No. 4, April, 1980
- Hickman, F.R., Killin, J.L., Land, L., Mulhall, T., Porter, D., Taylor, R.M., *Analysis for Knowledge-Based Systems, A Practical Guide to the KADS Methodology*, Ellis Horwood Books in Information Technology, John Wiley & Sons, New York, 1989
- Hofstede, G.H., *Culture's Consequences, International Differences in Work-Related Values*, London, Sage, 1980
- Hogarth, R.M., *Judgment and Choice*, John Wiley & Sons, New York, 1980
- Holstrum, G.W., Mock, T.J., West, R.N., *Information System Changes in the 1990's: Some Auditing Issues and Implications*, paper presented at the Rijksuniversiteit Limburg, 1990
- Hopwood, A.G., An Empirical Study of the Role of Accounting Data in Performance Evaluation, *Journal of Accounting Research (Supplement)*, 1972
- Hopwood, A.G., Bromwich, M., Shaw, J. (eds.), *Auditing Research: Issues and Opportunities*, Pitman Publishing, 1982
- House, R.J., Howell, J.M., Shamir, B., Smith, B., Spangler, W.D., *Charismatic Leadership: A 1990 Theory and Seven Empirical Tests*, paper presented at the Methodological Workshop of the Economic and Social Institute, Vrije Universiteit, Amsterdam, 1990
- James, W., *The Principles of Psychology*, New York. Holt. 1890
- Jenkins, B., Perry, R., Cooke, P., *An Audit Approach to Computers*, Coopers and Lybrand, The Institute of Chartered Accountants in England and Wales, 1986
- Jenssen, M.C., Meckling, W.H., Theory of the Firm: Managerial Behavior, Agency Cost and Ownership Structure, *Journal of Financial Economics*, 3, 1976

- Johnson, A.P., *Auditing Judgment, A Book of Cases*, Irwin, Homewood, Illinois, 1980
- Johnson, P.E., Jamal, K., Berryman, R.G., Audit Judgment Research, *Accounting, Organizations and Society*, Vol. 14, Nos. 1/2, 1989
- Johnson, P.E., The Expert Mind: A New Challenge for the Information Scientist, in: Bemelmans, M.A. (ed.), *Beyond Productivity: Information Systems Development for Organizational Effectiveness*, Elsevier Science Publishers, The Netherlands, 1984
- Joyce, E.J., Expert Judgment in Audit Program Planning: an Empirical Study, *Studies on Human Information Processing in Accounting, Supplement to the Journal of Accounting Research*, 14, pp. 29-60, 1976
- Joyce, E.J., Biddle, G.C., Anchoring and Adjustment in Probabilistic Inference in Auditing, *Journal of Accounting Research*, Spring 1981a
- Joyce, E.J., Biddle, G.C., Are Auditors' Judgments Sufficiently Regressive?, *Journal of Accounting Research*, Vol. 19, No. 2, Autumn 1981b
- Joyce, E.J., Libby, R., Behavioral Studies of Audit Decision Making, *Journal of Accounting Literature*, Vol. 1, 1982
- Jung, C.G., *Psychological Types*, Routledge & Kegan Paul Ltd., London, 1921
- Kachelmeier, S.J., Messier, W.F., An Investigation of the Influence of a Nonstatistical Decision Aid on Auditor Sample Size Decisions, *The Accounting Review*, Vol. 65, No. 1, January, 1990
- Kahneman, D., Tversky, A., The Psychology of Preferences, *Scientific American*, January, 1982
- Kahneman, D., Tversky, A., Subjective Probability: a Judgment of Representativeness, *Cognitive Psychology*, July, 1972
- Kaplan, R.S., *Advanced Management Accounting*, Prentice-Hall International, London, 1982
- Keen, P.G.K., Bronsema, G.L., *Cognitive Style Research: a Perspective for Integration*, unpublished paper, Sloan School of Management, 1981
- Keen, P.G.K., Scott-Morton, M., *Decision Support Systems: An Organizational Perspective*, Addison-Wesley, 1978
- Kerin, R.A., Slovic, J.W., Decision-making Style and Acquisition of Information: Further Exploration of the Myers-Briggs Type Indicator, *Psychological Reports*, 49, 1981
- Kinney, W.R., A Decision-Theory Approach to the Sampling Problem in Auditing, *Journal of Accounting Research*, Spring, 1975
- Kinney, W.R., Audit Technology and Preferences for Auditing Standards, *Journal of Accounting and Economics*, 8, March, 1986
- Kinney, W.R., Attestation Research Opportunities: 1987, *Contemporary Accounting Research*, Vol. 4, 1988
- Kinney, W.R., Uecker, W.C., Mitigating the Consequences of Anchoring in Auditor Judgment, *The Accounting Review*, January, 1982



Klersey, G.F., Mock, Th.J., Verbal Protocol Research in Auditing, *Accounting, Organizations and Society*, Vol. 14, Nos. 1/2, 1989

Klersey, G.F., Mock, T.J., Verbal Protocol Research in Auditing, *Accounting, Organizations and Society*, Vol. 14, 1989

Kolodner, J.L., Towards an Understanding of the Role of Experience in the Evolution from Novice to Expert, *International Journal of Man-Machine Studies*, 19, November, 1983

Krogstad, J.L., Ettenson, R.T., Shanteau, J., Context and Experience in Auditors' Materiality Judgments, *Auditing: a Journal of Practice and Theory*, Vol. 4, No. 1, Fall, 1984

Larcker, D.F., Lessig, V.P., An Examination of the linear and Retrospective Process Tracing Approaches to Judgment Modeling, *The Accounting Review*, Vol. 58, No. 1, January, 1983

Lembre, E. de, Dries, R., Risico-analyse in Auditing, *Accountancy en Bedrijfskunde*, March, 1987

Leslie, D.A., *Materiality. The Concept and its Application to Auditing*, CICA, 1985

Lewis, B.L., Expert Judgment in Auditing: An Expected Utility Approach, *Journal of Accounting Research*, Vol. 18, No. 2, Autumn, 1980

Libby, R., Frederick, D.M., Experience and the Ability to Explain Audit Findings, *Journal of Accounting Research*, Vol. 28, No. 2, 1990

Libby, R., Lewis, B.L., Human Information Processing in Accounting: The State of the Art in 1982, *Accounting, Organizations and Society*, 3, 1982

Libby, R., Lewis, B.L., Human Information Processing Research in Accounting: The State of the Art, *Accounting, Organizations and Society*, 2, 1977

Libby, R., *Accounting and Human Information Processing*, Prentice-Hall, Englewood Cliffs, 1981

Libby, R., Behavioral Research in Accounting, in Jensen, D.L.(ed.), *Accounting Dissertations, Research, Design and Implementation*, College of Administrative Science, Ohio State University, Columbus, Ohio, 1982

Libby, R., Availability and the Generation of Hypotheses in Analytical Review, *Journal of Accounting Research*, Autumn, 1985

Libby, R., *Accounting and Human Information Processing*, Prentice-Hall, Englewood Cliffs, 1981

Limperg Instituut, *Risico's en Oordeelsvorming in de Accountantspraktijk*, Limperg Institute, Amsterdam, 1990

Loebbecke, J.K., Ward, B., Discussion of Materiality Allocation in Audit Planning: A Feasibility Study, *Journal of Accounting Research*, Vol. 17, Supplement, 1979

Lusk, E.J., A Test of Differential Performance Peaking for a Disembedding Task, *Journal of Accounting Research*, Spring, 1979

Lyon, D., Slovic, P. Dominance of Accuracy Information and Neglect of Base Rates in Probability Estimation, *Acta Psychologica*, 1976

- MacDonald, A.P., Revised Scale for Ambiguity Tolerance: Reliability and Validity, *Psychological Reports*, 26, 1970
- Macintosh, N.B., *The Social Software of Accounting and Information Systems*, John Wiley & Sons, 1985
- Mautz, R.K., Sharaf, H.A., *The Philosophy of Auditing*, American Accounting Association, Monograph No. 6, 11th ed., 1982
- Mautz, R.K., Evidence, Judgment and the Auditor's opinion, *Journal of Accountancy*, April 1959
- McCarty, L.T., Reflections on TAXMAN: An Experiment in Artificial Intelligence and Legal Reasoning, *Harvard Law Review*, Vol. 90, No. 5, 1977
- Meservy, R.D., Bailey, A.D. Jr., Johnson, P.E., Internal Control Evaluation: A Computational Model of the Review Process, *Auditing: A Journal of Practice and Theory*, Fall, 1986
- Meservy, R.D., *Auditing Internal Controls: a Computational Model of the Review Process*, Ph.D thesis, University of Minnesota, August, 1985
- Messier, W., The Effect of Experience and Firm Type on Materiality/Disclosure Judgments, *Journal of Accounting Research*, Autumn, 1983
- Miller, G.A., the Magical Number Seven, Plus-or-Minus Two: Some Limits on out Capacity for Processing Information, *Psychological Review*, 63, 1956
- Mock, T.J., Estrin, T.L., Vasarhelyi, M.A., Learning Patterns, Decision Approach, and Value of Information, *Journal of Accounting Research*, Spring, 1972
- Mock, T.J., Turner, J.L., The Effect of Changes in Internal Controls on Audit Programs, in Burns, T.J. (ed.), *Behavioral Experiments in Accounting II*, College of Administrative Science, Ohio State University, Columbus, Ohio, 1979
- Mock, T.J., Turner, J.L., *Internal Accounting Control Evaluation and Auditor Judgment*, Auditing Monograph no. 3, AICPA, New York, 1981
- Mock, T.J., Vertinsky, I., *Risk Assessment in Accounting and Auditing*, The Canadian Certified General Accountants' Research Foundation, 1985
- Mock, T.J., Washington, M.T., Risk Concepts and Risk Assessment in Auditing, *Advances in Accounting*, Supplement 1, 1989
- Mock, T.J., Wright, A., Washington, M., Krishnamoorthy, G., *Auditors' Probabilistic Reasoning in a Multi-Stage Risk Assessment Task*, forthcoming, 1993
- Mock, T.J., Wright, A., *An Exploratory Study of Auditor Evidential Planning Judgments*, revised version of a paper presented at the 1991 Annual Congress of the European Accounting Association, Maastricht, The Netherlands, 1992
- Moeckel, C., The Effect of Experience on Auditors' Memory Errors, *Journal of Accounting Research*, Vol. 28, No. 2, 1990
- Moeckel, C., Two Factors Affecting an Auditor's Ability to Integrate Audit Evidence, *Contemporary Accounting Research*, Vol. 8, Nr. 1, Fall, 1991

- Moriarty, S., Barron, F.H., A Judgment-Based Definition of Materiality, Studies on Auditing, Selections from the 'Research Opportunities in Auditing'-Program, *Supplement to Journal of Accounting Research*, 1979
- Moriarty, S., Barron, F.H., Modeling the Materiality Judgments of Audit Partners, *Journal of Accounting Research*, Autumn, 1976
- Morris, M.H., Nichols, W.D., Consistency Exceptions: Materiality Judgments and Audit Firm Structure, *The Accounting Review*, Vol. LXIII, No. 2, April, 1988
- Müller, G.E., Zur Analyse der Gedachtnistätigkeit und Vorstellungsverlaufes: Teil I, *Zeitschrift für Psychologie*, 5, 1911
- Murphy, D.S., Expert System Use and the Development of Expertise in Auditing: a Preliminary Investigation, *Journal of Information Systems*, Fall, 1990
- Murray, J.B., Review of Research on the Myers-Briggs Type Indicator, *Perceptual and Motor Skills*, 70, 1990
- Neale, I.M., First Generation Expert Systems: a Review of Knowledge Acquisition Methodologies, *The Knowledge Engineering Review*, Vol. 3, No. 2, June, 1988
- Newell, A., The Knowledge Level, *Artificial Intelligence Journal*, Vol. 19, No. 2, 1982
- Newell, A., Shaw, J.C., Simon, H.A., Elements of a Theory of Human Problem Solving, *Psychological Review*, 65, 1958
- Newell, A., Simon, H., *Human Problem Solving*, Prentice-Hall, Englewood Cliffs, 1972
- Newton, L.K., The Risk Factor in Materiality Decisions, *The Accounting Review*, January, 1977
- Nisbett, R.E., Wilson, T.D., Telling More than We Can Know: Verbal Reports on Mental Processes, *Psychological Review*, 84, 1977
- Nutt, P.C., Strategic Decisions Made by Top Executives and Middle Managers with Data and Process Dominant Styles, *Journal of Management Studies*, 27, March, 1990
- Pany, K., Whittington, O.R., *Auditing*, Irwin, 1994
- Payne, J.W., Braunstein, M.L., Carroll, J.S., Exploring Pre-decisional Behavior: An Alternative Approach to Decision Research, *Organizational Behavior and Human Performance*, 22, February, 1978
- Parker, L.D., Ferris, K.R., Otley, D.T., *Accounting for the Human Factor*, Prentice Hall, 1989
- Peters, J.M., *Assessing Inherent Risk in an External Audit: a Computational Model*, working paper, University of Pittsburgh Graduate School of Business, January, 1987
- Peterson, C., Beach, L., Man as an Intuitive Statistician, *Psychological Bulletin*, July, 1967
- Phelan, J.G., A Replication of a Study on the Effects of Attempts to Verbalize on the Process of Concept Attainment, *Journal of Psychology*, 59, 1965

- Pincus, K.V., Auditor Individual Differences and Fairness of Presentation Judgments, *Auditing: A Journal of Practice and Theory*, Vol. 9, No. 3, Fall, 1990
- Pincus, K.V., Audit Judgment Confidence, *Behavioral Research in Accounting*, Vol. 3, 1991
- Poel, J.H.R. van de, Vaassen, E.H.J., *Audit Judgment Research, Accountantscontrole en Menselijke Beslissingsprocessen*, Research Memorandum, Rijksuniversiteit Limburg, RM 88-033, December, 1988
- Poel, J.H.R. van de, *Als de Teken en Bedriegen*, inaugural lecture, delivered in a shorter version on taking up the chair of professor in business economics at the University of Limburg, Wolters-Noordhoff, Groningen, 1988
- Poel, J.H.R. van de, *Judgment and Control, Individual and Organizational Aspects of Performance Evaluation*, Wolters-Noordhoff, Groningen, 1986
- Poel, J.H.R. van de, Door de Hond of de Kat, Paper, forthcoming 1994
- Rosenberg, M., *The Logic of Survey Analysis*, Basic Books, New York, 1968
- Ross, S.A., The Economic Theory of Agency: The Principal's Problem, *American Economic Review*, Vol. 72, No. 3, 1973
- Ruble, T.L., Cosier, R.A., Effects of Cognitive Styles and Decision Setting on Performance, *Organizational Behavior and Human Decision Processes*, 46, 1990
- Russo, J.E., Comments on Behavioral and Economic Approaches to Studying Market Behavior, in Mitchell, A.A. (ed.), *The Effect of Information on Consumer and Market Behavior*, American Market Association, Chicago, 1978
- Schweiger, D.M., Jago, A.G., Problem-solving Styles and Participative Decision Making, *Psychological Reports*, 50, 1982
- Scott, W.R., Discussion of Materiality Allocation in Audit Planning: A Feasibility Study, *Journal of Accounting Research*, Vol. 17, Supplement, 1979
- Scott, W.R., The State of the Art of Academic Research in Auditing, *Journal of Accounting Literature*, Vol. 3, 1984
- Seiler, R.E., Bartlett, R.W., Personality Variables as Predictors of Budget System Characteristics, *Accounting, Organizations and Society*, Vol. 7, No. 4, 1982
- Shields, M.D., Solomon, I., Waller, W.S., Effects of Alternative Sample Space Representations on the Accuracy of Auditors' Uncertainty Judgments, *Accounting, Organizations and Society*, Vol. 12, No. 4, 1987
- Siegel, S., Castellan, N.J., *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, statistics series, 1988
- Simon, H.E., How Big is a Chunk?, *Science*, 1974
- Simons, W.H.G., *De Regulering van Inferenties door de Kennis van de Lezer*, Proefschrift, Katholieke Universiteit Nijmegen, 1993

- Slovic, P., Fischhoff, B., Lichtenstein, S., Behavioral Decision Theory, *Annual Review of Psychology*, 28, 1977
- Slovic, P., Lichtenstein, S., Comparison of Bayesian and Regression Approaches to the Study of Information Processing in *Judgment, Organizational Behavior and Human Performance*, June, 1971
- Shpilberg, D., Graham, L.E., Developing ExpertTAX: an Expert System for Corporate Tax Accrual and Planning, *Auditing; a Journal of Practice and Theory*, Fall, 1986
- Simon, H., *The New Science of Management*, Harper and Row, 1960
- Snieszak, J.A., Henry, R.A., Accuracy and Confidence in Group Judgment, *Organizational Behavior and Human Decision Processes*, Vol. 43, 1989
- Soeters, J., Schreuder, H., The Interaction Between National and Organizational Cultures in Accounting Firms, *Accounting, Organizations and Society*, Vol. 13, No. 1, 1988
- Sorter, G.H., An Events Approach to Basic Accounting Theory, *The Accounting Review*, January, 1969
- Steels, L., *Kennissystemen*, Addison-Wesley Publishing Company, 1992
- Steinbart, P.J., The Construction of a Rule Based Expert System as a Method for Studying Materiality Judgments, *The Accounting Review*, Vol. LXII, No.1, 1987
- Swagerman, J., META, een Expert Systeem dat de Accountant Helpt bij de Beoordeling van Expertsystemen, *Maandblad voor Accountancy en Bedrijfseconomie*, June, 1989
- Tabor, R.H., Internal Control Evaluations and Audit Program Revisions: Some Additional Evidence, *Journal of Accounting Research*, Vol. 21, No. 1, 1983
- Titchener, E.B., The Schema of Introspection, *American Journal of Psychology*, 23, 1912
- Turing, A.M., On Computable Numbers, with an Application to the Entscheidungsproblem, in: *Proceedings of the London Mathematics Society*, London, 1936
- Tversky, A., Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science*, September, 1974
- Uecker, W., Kinney, W.R., Judgmental Evaluation of Sample Results: a Study of the Type and Severity of Errors made by Practicing CPA's, *Accounting, Organizations and Society*, 2, 1977
- Vaassen, E.H.J., Hayes, R.S., *Cognitive Characteristics and Information Search of Senior Auditors in the Netherlands*, paper presented at the 13th Annual Congress of the European Accounting Association, 1990
- Vaassen, E.H.J., Hayes, R.S., Baker, C.R., *An Investigation of Auditor Cognitive Styles*, paper presented at the American Accounting Association 1992 Annual Meeting, August, 1992
- Vaassen, E.H.J., Materialiteitsnormen bij de Accountantscontrole, *Maandblad voor Accountancy en Bedrijfseconomie*, No. 4, April, 1991
- Vaassen, E.H.J., *The Development of a Rule Based Expert System for Planning Stage Materiality Judgments*, paper presented at the 14th Annual Congress of the European Accounting Association, 1991

- Vasarhelyi, M., Man-Machine Planning Systems: a Cognitive Style Examination of Interactive Decision Making, *Journal of Accounting Research*, Spring, 1977
- Vasarhelyi, M., Halper, F.B., The Continuous Audit of Online Systems, *Auditing: a Journal of Practice and Theory*, Spring, 1991
- Vermunt, J.D.H.M., Lodewijks, J.G.L.C., Simons, P.R.J., Hardop-denken als onderzoeksmethode naar regulatieprocessen bij tekstbestudering, *Tijdschrift voor Onderwijsresearch*, 11, nr. 4, 1986
- Verplanck, W.S., Unaware of Where's Awareness: Some Verbal Operants-notates, in: Eriksen (ed.), *Behavior and Awareness, A Symposium of Research and Interpretation*, Durham, Duke University Press, 1962
- Wallage, Ph., *Methodiek en Mate van Structuur: een Beschouwing over het Proces van Accountantscontrole*, dissertation, University of Amsterdam, the Netherlands, June 1991
- Waller, W.S., Felix, W.L., The Auditor and Learning from Experience: Some Conjectures, *Accounting Organizations and Society*, Vol. 9, No. 3/4, 1984
- Waller, W.S., Felix, W.L., Auditors' Covariation Judgments, *The Accounting Review*, Vol. 62, No. 2, April 1987
- Waller, W., Jiambalvo, J., The Use of Normative Models in Human Information Processing Research in Accounting, *Journal of Accounting Literature*, Vol. 3, 1984
- Ward, B.H., An Investigation of The Materiality Construct in Auditing, *Journal of Accounting Research*, Spring, 1976
- Waterman, D.A., Newell, A., Protocol Analysis as a Task for Artificial Intelligence, *Artificial Intelligence*, 2, 1971
- Waterman, D.A., *A Guide to Expert Systems*, Addison-Wesley Publishing Company, 1986
- Watson, J.B., Psychology as the Behaviorist Views It, *Psychological Review*, 20, 1913
- Watson, J.B., Is Thinking Merely the Action of Language mechanisms?, *British Journal of Psychology*, 11, 1920
- Watts, R.L., Zimmerman, J.L., Towards a Positive Theory of the Determination of Accounting Standards, *The Accounting Review*, January, 1978
- Watts, R.L., Zimmerman, J.L., Agency Problems, Auditing and the Theory of the Firm: some Evidence, *Journal of Law and Economics*, 26, 1, 1983
- Watts, R.L., Zimmerman, J.L., Positive Accounting Theory: A Ten Year Perspective, *The Accounting Review*, Vol. 65, No. 1, January, 1990
- Weber, R., Auditor Decision Making on Overall System Reliability: Accuracy, Consensus and the Usefulness of a Simulation Decision Aid, *Journal of Accounting Research*, Autumn, 1978
- Weber, R., Some Characteristics of the Free Recall of Computer Controls by EDP Auditors, *Journal of Accounting Research*, Spring, 1980

- White, C.E. Jr., *The Maturing of Expert Systems in Accounting and Auditing*, paper presented to: 'Using Expert Systems by Accountants and Auditors'; an International Conference Organized by The Committee on Auditing and EDP of NIVRA, Amsterdam, The Netherlands, September, 1988
- Williams, J.D., *Risk Assessment: An Investigation of Auditors' Decision Processes*, paper presented at the Audit Judgment Symposium of the University of Southern California, 1989
- Williamson, O.E., *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, New York, 1975
- Winkler, R.L., The Assessment of Prior Distributions in Bayesian Analysis, *Journal of the American Statistical Association*, Vol. 62, 1967
- Witkin, H.A., Individual Differences in Ease of Perception of Embedded Figures, *Journal of Personality*, 19, 1950
- Woodworth, R.S., *Experimental Psychology*, New York, Holt, Rinehart & Winston, 1938
- Woolf, E., *Auditing Today*, Prentice/Hall, Englewood Cliffs, third ed., 1986
- Wouters, L., Jong. T. de, Hardop Denken tijdens Tekstbestudering, *Tijdschrift voor Onderwijsresearch*, 7, nr. 2, 1982
- Wright, W.F., Audit Judgment, Consensus and Experience, in: Ferris, K.R., (ed.), *Behavioral Accounting Research: a Critical Analysis*, Columbus, Century VII Publishing Co., 1988
- Wright, G., *Behavioural Decision Theory, An Introduction*, Penguin Books, 1984
- Wundt, W., *Grundriss der Psychology*, Alfred Kröner, Leipzig, 1914





# Nederlandse samenvatting

(Summary in Dutch)

## *Beslissingsprocessen van accountants bij oordeelsvorming inzake materialiteit in de planningfase van de accountantscontrole*

Accountantscontrole is erop gericht fouten van een bepaalde omvang in een financiële verantwoording te vinden en te (laten) corrigeren. De omvang van fouten die de accountant wenst te vinden wordt door de accountant in de planningfase van de accountantscontrole vastgesteld. De vraag die de accountant zich daarbij stelt is: zal een rationele gebruiker van die verantwoording een andere beslissing nemen dan die welke hij genomen zou hebben als een bepaalde fout niet in de financiële verantwoording aanwezig zou zijn. De omvang van de fouten die door de accountant geaccepteerd kunnen worden, wordt bepaald door de materialiteitsgrens.

In elke accountantscontrole worden door de accountant beslissingen genomen waarvan de uitkomsten niet met volledige zekerheid vaststaan. Om tot een accountantsverklaring te komen zullen accountants derhalve vakkundige oordeelsvorming toepassen. Toepassing van vakkundige oordeelsvorming vereist een bepaalde mate van expertise. In deze studie wordt expertise gedefinieerd als het positieve effect op de kwaliteit van beslissingen dat uitgaat van aangeboren eigenschappen die een optimale afstemming tussen persoon en taak mogelijk maken, werkervaring en opleiding. Het is zinvol om de condities waaronder expertise aangekweekt kan worden bij accountants en aankomend accountants te kennen omdat accountantskantoren alsdan sneller de beschikking kunnen krijgen over experts in accountancy die vakkundige oordeelsvorming optimaal kunnen toepassen. Gegeven de gehanteerde definitie van expertise zijn er drie categorieën van condities waarvan kennis noodzakelijk is: psychologisch profiel, werkervaring (inclusief accountantskantoor waar men werkzaam is) en opleiding. Het aantal jaren werkervaring, het accountantskantoor waaraan de accountant verbonden is en de aard van de opleiding tot registeraccountant worden in deze studie samengevat onder de noemer: demografische kenmerken.

In deze studie wordt een algemeen model van vakkundige oordeelsvorming in de accountantscontrole geïntroduceerd. Uitgaande van dit model worden de volgende drie problemen onderzocht:

1. Hebben accountants een bepaald overheersend psychologisch profiel en is er een relatie tussen dit psychologisch profiel en het aantal jaren werkervaring bij een accountantskantoor en de opleiding tot registeraccountant die gevolgd is?
2. Is er een relatie tussen psychologisch profiel en demografische kenmerken van accountants en de wijze waarop zij problemen oplossen?
3. Welke factoren bepalen de hoogte van de materialiteit zoals vastgesteld in de planningfase van de accountantscontrole?

Om deze vragen te beantwoorden werd een case ontwikkeld, waarin de opdracht was een uitspraak te doen over de materialiteitsgrens op basis van financiële en

niet-financiële informatie. In de case werd tevens de mogelijkheid opengelaten om de materialiteitsvraagstelling niet te behandelen maar een advies inzake een overnameprijs uit te brengen. De keuze voor één van beide richtingen mocht door de proefpersoon zelfstandig gemaakt worden. De case werd via de computer aan 25 registeraccountants, verspreid over vier grote accountantskantoren, met gemiddeld 21,4 jaren ervaring in de accountantscontrolepraktijk, gepresenteerd. Door de case via de computer te presenteren werd het mogelijk een gedetailleerde registratie van informatie-opvraag activiteiten door de proefpersonen bij te houden. Voorts werd aan de proefpersonen gevraagd tijdens het oplossen van de case hardop te denken. Van de uitspraken die de proefpersonen tijdens het probleemoplossen deden werden audio-opnamen gemaakt en van de overige activiteiten tijdens het probleemoplossen werden video-opnamen gemaakt, dit om een zo volledig en nauwkeurig mogelijke registratie van uiterlijke kenmerken van het probleemoplossingproces van de proefpersonen te verkrijgen. Als de proefpersonen te kennen gaven dat ze de case dachten te hebben opgelost werd een persoonlijkheidstest (MacDonald's test van tolerantie voor ambiguïteit met behulp waarvan gemeten wordt in welke mate mensen een voorkeur hebben voor verschijnselen die niet eenduidig uit te leggen zijn) en een cognitieve stijl-test (Myers-Briggs type indicator naar een typologie van Jung met behulp waarvan onder andere getest wordt op welke wijze mensen nieuwe informatie in zich opnemen en die informatie verwerken) aan hen voorgelegd. Nadat ze deze tests hadden ingevuld werd de vraag gesteld wat ze zich nog konden herinneren van hun gedachten tijdens het oplossen van de case. Van de uitspraken die aldus werden gedaan, werden eveneens audio-opnamen gemaakt. Op basis van de letterlijke transcripties van de uitspraken van de proefpersonen tijdens en na hun beslissingsprocessen en op basis van controlehandboeken en tekstboeken controleleer werd een prototype expertsysteem gebouwd dat als enige functie had het in een tweede onderzoeksrunde vergemakkelijken van het verkrijgen van additionele informatie aangaande de beslissingsprocessen van de proefpersonen. Eveneens op basis van de letterlijke transcripties van de uitspraken van de proefpersonen tijdens en na hun beslissingsprocessen werd volgens de methode van verbale protocolanalyse een codering gemaakt van de uitspraken met het doel statistische analyse mogelijk te maken. In de tweede ronde werd aan 17 proefpersonen gevraagd om op basis van minimaal één door henzelf ingebrachte case aan te geven of ze het eens danwel oneens waren met conclusies die het prototype expertsysteem trok. Dit proces leidde tot een verfijning van het prototype expertsysteem. Het aldus ontstane verfijnde prototype werd tenslotte voorgelegd aan de resterende proefpersonen met de opdracht de vragen die het prototype expertsysteem stelde te beantwoorden aan de hand van een tweetal gestructureerde cases. Door een dergelijk iteratief proces te volgen werd informatie verkregen over factoren die de planningmaterialiteit kunnen beïnvloeden.

Onderzoeksvraag 1 (psychologisch profiel en demografische kenmerken):

De eerste onderzoeksvraag leidde tot de conclusie dat de meeste accountants een

hoge tolerantie voor ambiguïteit hebben doch dat er géén overheersende cognitieve stijl-categorie is. Dit laatste betekent dat accountants zowel een waarnemende als een intuïtieve cognitieve stijl voor informatie acquisitie kunnen hebben en dat ze zowel een denkende als een gevoelsmatige cognitieve stijl voor informatie-verwerking kunnen hebben. Voorts bleek dat er géén relatie tussen het psychologisch profiel en het aantal jaren werkervaring is, wat tot de conclusie moet leiden dat het psychologisch profiel - conform de definitie - niet wijzigt in de loop der tijd, danwel dat de groep proefpersonen een dermate hoog ervaringsniveau heeft dat er nauwelijks verschillen bestaan tussen de relatief onervaren proefpersonen en de relatief ervaren proefpersonen. Verder bleek dat er géén verschil is in psychologisch profiel tussen accountants met een NIVRA-opleiding en accountants met een academische opleiding. Tenslotte bleek dat de proefpersonen die verbonden zijn aan het accountantskantoor in het onderzoek met de minst gestructureerde controle-aanpak een hogere tolerantie voor ambiguïteit en een meer gevoelsmatige cognitieve stijl (als tegenpool van een denkende cognitieve stijl) hebben dan de proefpersonen die verbonden zijn aan de accountantskantoren met de meer gestructureerde controle-aanpak.

Onderzoeksvraag 2 (psychologisch profiel, demografische kenmerken en de wijze van probleemoplossen):

De tweede onderzoeksvraag kan worden opgesplitst in drie deelvragen. De eerste deelvraag betreft de relatie tussen de voorkeuren van accountants voor informatie-opvraag, het gebruikmaken van bestaande kennis, de oordeelsvorming en het nemen van beslissingen enerzijds en de persoonlijke kenmerken van accountants anderzijds. De tweede deelvraag betreft de relatie tussen de hoeveelheid herinnerde informatie en de volledigheid van informatie-opvraag enerzijds en de persoonlijke kenmerken van accountants anderzijds. De derde deelvraag betreft de relatie tussen de mate waarin accountants het eens zijn over hun daadwerkelijke informatie-opvraag en de tijd die aan die opgevraagde informatie besteed wordt enerzijds en hun persoonlijke kenmerken anderzijds.

Onderzoeksvraag 2a (de relatie tussen de voorkeuren van accountants voor informatie-opvraag, het gebruikmaken van bestaande kennis, de oordeelsvorming en het nemen van beslissingen enerzijds en de persoonlijke kenmerken van accountants anderzijds):

De eerste deelvraag bij onderzoeksvraag twee leidde tot de conclusie dat de relatief ervaren accountants minder informatie wensen op te vragen, maar meer gebruik wensen te maken van hun bestaande kennis dan de relatief onervaren accountants. Voorts bleek dat de accountants met een NIVRA-opleiding meer informatie wensen op te vragen dan de accountants met een academische opleiding. Verder bleek dat de accountants van de accountantskantoren met een verhoudingsgewijs ongestructureerde controle-aanpak minder vaak tot een beslissing (i.c. inzake materialiteit) komen, maar dat binnen deze groep de accountants van het

accountantskantoor met de meest ongestructureerde controle-aanpak meer gebruik wensen te maken van hun bestaande kennis en minder oordeelsvorming toepassen. Met betrekking tot cognitieve stijl bleek dat de accountants met een waarnemende cognitieve stijl meer informatie wensen op te vragen dan accountants met een intuïtieve cognitieve stijl en dat de accountants met een gevoelsmatige cognitieve stijl meer gebruik wensen te maken van hun bestaande kennis, een kleiner aantal relevante uitspraken doen en minder tijd nodig hebben om tot een beslissing te komen dan de accountants met een denkende cognitieve stijl. Tenslotte, met betrekking tot de persoonlijkheid, zoals gemeten middels de test waarin de tolerantie voor ambiguïteit van personen wordt vastgesteld, bleek dat er géén voorkeuren voor het opvragen van informatie, het gebruikmaken van de bestaande kennis, het combineren van nieuwe informatie met de bestaande kennis danwel het nemen van een beslissing werden kenbaar gemaakt door de accountants met een hoge c.q. een lage tolerantie voor ambiguïteit. Tevens bleek dat er géén voorkeuren voor het expliciet kiezen voor één van de twee richtingen in de case werden kenbaar gemaakt door de accountants met een hoge c.q. een lage tolerantie voor ambiguïteit.

Onderzoeksvraag 2b (de relatie tussen de hoeveelheid herinnerde informatie en de volledigheid van informatie-opvraag enerzijds en de persoonlijke kenmerken van accountants anderzijds):

De tweede deelvraag bij onderzoeksvraag twee leidde tot de conclusie dat er een interactie-effect tussen het aantal jaren ervaring en volledigheid van informatie-opvraag bestaat in die zin dat voor de relatief ervaren accountants geldt dat veel informatie-opvraag tot een grotere hoeveelheid herinnerde informatie leidt, en dat voor relatief onervaren accountants geldt dat veel-informatie-opvraag tot een kleinere hoeveelheid herinnerde informatie leidt. Er zijn géén effecten van de genoten opleiding of het accountantskantoor op de hoeveelheid herinnerde informatie geconstateerd. Voorts bleek dat de accountants met een waarnemende cognitieve stijl een grotere hoeveelheid herinnerde informatie hebben dan de accountants met een intuïtieve cognitieve stijl, doch dat géén verschil betreffende de hoeveelheid herinnerde informatie bestaat tussen de accountants met een gevoelsmatige cognitieve stijl en accountants met een denkende cognitieve stijl. Tenslotte bleek dat de accountants met een hoge tolerantie voor ambiguïteit een grotere hoeveelheid herinnerde informatie hebben dan de accountants met een lage tolerantie voor ambiguïteit.

Onderzoeksvraag 2c (de relatie tussen de mate waarin accountants het eens zijn over hun daadwerkelijke informatie-opvraag en de tijd die aan die opgevraagde informatie besteed wordt enerzijds en hun persoonlijke kenmerken anderzijds):

De derde deelvraag bij onderzoeksvraag twee leidde tot de conclusie dat er onder de relatief onervaren accountants meer overeenstemming bestaat inzake de tijd die aan bepaalde informatie besteed zal worden dan onder de relatief ervaren

accountants. Er zijn géén effecten van opleiding of accountantskantoor op de mate van overeenstemming inzake de volgorde van informatie-opvraag danwel de benodigde tijd om informatie te lezen en te becommentariëren geconstateerd. Verder bleek dat onder de accountants met een waarnemende cognitieve stijl meer overeenstemming bestaat aangaande de volgorde van hun informatie-opvraag dan onder de accountants met een intuïtieve cognitieve stijl en dat onder de accountants met een denkende cognitieve stijl meer overeenstemming bestaat aangaande het sequentieel opvragen van informatie dan onder de accountants met een gevoelsmatige cognitieve stijl. Tenslotte bleek dat onder de accountants met een hoge tolerantie voor ambiguïteit meer overeenstemming bestaat aangaande het sequentieel (d.w.z. in de volgorde zoals weergegeven in de inhoudsopgave van de op te vragen informatie) opvragen van informatie en dat onder de accountants met een lage tolerantie voor ambiguïteit meer overeenstemming bestaat aangaande de volgorde van informatie-opvraag van de groep accountants met een lage tolerantie voor ambiguïteit dan onder de accountants met een hoge tolerantie voor ambiguïteit.

Onderzoeksvraag 3 (materialiteitsbepalende factoren):

De derde onderzoeksvraag leidde tot de conclusie dat de materialiteit in de planningfase van de accountantscontrole bepaald wordt als het produkt van een materialiteitsbasis, een daarbij behorend percentage en een vermenigvuldigingsfactor. De materialiteitsbasis en het percentage dat daarbij behoort zijn gebaseerd op financiële grootheden. De vermenigvuldigingsfactor wordt echter bepaald door accountantscontrole-risicofactoren, wat verwerpelijk is vanuit een statistisch standpunt, en door theoretisch wél aanvaardbare factoren (*echte* materialiteitsbepalende factoren). Tot de accountantscontrole-risicofactoren behoren bijvoorbeeld: tendenties tot het flatteren danwel deflaten van de winst, de branche van de onderneming en de kwaliteit van de interne/administratieve organisatie. Tot de echte materialiteitsbepalende factoren behoren bijvoorbeeld: de invloed van aandeelhouders, het al dan niet aan de beurs genoteerd zijn en de aard en omvang van fouten die niet in een volgende periode gecorrigeerd kunnen worden. Bij het vaststellen van materialiteitsbepalende factoren bleek dat er weinig overeenstemming bestaat aangaande de hoogte van de materialiteit in de gepresenteerde case en dat accountants het onderscheid tussen accountantscontrole-risico en materialiteit onvoldoende kennen en diens gevolge ook niet maken. Dit is overigens verklaarbaar uit het feit dat accountants géén terugkoppeling over genomen materialiteitsbeslissingen krijgen vanuit het maatschappelijk verkeer en aldus niet kunnen leren van gemaakte fouten inzake materialiteitsvaststelling.

Middels deze studie is een aantal elementen van het algemeen model van vakkundige oordeelsvorming in de accountantscontrole onderzocht. De belangrijkste conclusie is dat er significante relaties onderkend kunnen worden tussen het

