The effects of feedback on audit judgement performance under different levels of task complexity

Author:
Leung, Wing Yin Patrick

Publication Date:
2002

DOI:
https://doi.org/10.26190/unsworks/5213

THE EFFECTS OF FEEDBACK ON AUDIT JUDGMENT PERFORMANCE UNDER DIFFERENT LEVELS OF TASK COMPLEXITY

A Dissertation in Fulfilment of Requirements for the Degree of Doctor of Philosophy

by

Wing Yin (Patrick) Leung

2002

SCHOOL OF ACCOUNTING

THE UNIVERSITY OF NEW SOUTH WALES

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

ABSTRACT

This thesis examines the relative effectiveness of different types of feedback in improving auditors' judgment performance for two audit tasks of different levels of task complexity. Specifically, judgment performance was examined across five types of feedback: no feedback (control group), outcome feedback, task properties feedback, cognitive feedback, and combined feedback (task properties and cognitive feedback) for two audit risk assessment tasks, which differ at two levels of task complexity. The high complexity task involved configural cue processing and the low complexity task did not.

The experiment involving 224 experienced auditors was a 5 (feedback treatments) × 2 (task complexity) × 2 (blocks) full factorial design. Feedback and task complexity were manipulated between subjects and block was a within subjects variable. To measure judgment performance, judgment accuracy was calculated as the mean absolute error or difference between each of the subjects 16 risk assessments in each block, and the judgments of a selected audit partner. Judgment achievement (r_a), and other lens model statistics, were also computed. To test for a transfer of learning, a second task (involving configural cue processing) was also completed.

Outcome feedback was effective in improving judgment performance in a low complexity task but not in a high complexity one. In line with previous studies, task properties feedback and combined feedback were found to be effective in improving judgment performance for both high and low complexity tasks. The thesis provided the first examination of cognitive feedback in the auditing literature. Cognitive feedback was found to improve judgment performance in a task of high complexity, but not in a low complexity one. The result helps to explain the general finding of the ineffectiveness of cognitive feedback in most previous studies that used simple generic tasks. One important finding is that self-insight moderates the effect of cognitive feedback in improving judgment performance. In addition to the practical implications for auditing, this finding has theoretical implications for both the psychology and auditing literatures.

The transferability of learning from one audit task to another is also investigated. It was found that subjects who completed a configural task initially, processed information more configurally on a subsequent configural task than did subjects who initially completed a non-configural task. It was also found that practice in a configural task coupled with receiving combined feedback led to a significantly better transfer of knowledge and judgment performance than no feedback or outcome feedback.

Task properties, cognitive, and combined feedback also led to significantly more auditors exhibiting configural cue processing, and auditors attributing significantly more weight to the configural cue combination in the second block of trials. Subjects did not exhibit such a change in the no feedback and outcome feedback treatments.

ACKNOWLEDGEMENT

To my supervisor, Scientia Professor Ken Trotman, my sincere gratitude for his insightful comments, guidance, encouragement, patience and continuous support throughout the life of this thesis.

This thesis has also benefited from the comments and discussion of many other people, and particularly, Cameron Hooper, Peter Luckett, Marcus O'Connor, and Ira Solomon.

I wish to thank the 236 practising auditors and the partners who participated in this experiment. Special thanks are extended to the following individuals who were involved in the pilot test: Kennedy Liu and Bobby Wong (Arthur Andersen); Kay Chan, Peter Cheung, and Tommy Wong (Deloitte, Touche and Tohmatsu); Melody Yeung (Ernst and Young); Louise Liu, Michelle Leung, and Magdalene Sun (KPMG); Raymond Kong, Kenny Tse, Carmen Wong and Carrie Yu (PricewaterhouseCoopers).

Thanks are also extended to Roger Lui and Brian Sum for their programming assistance, Monique Hui for her assistance in data processing, and Rebekah Solomon for her editing of the thesis. Many of my friends and colleagues have assisted in the project.

My sincere thanks to my family for their understanding and support throughout the life of this thesis.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

CHAPTER 4

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

Page

LIST OF TABLES, EXHIBITS, AND FIGURES

LIST OF TABLES, EXHIBITS, AND FIGURES (Continued)

CHAPTER 1

INTRODUCTION

1.1 Feedback and Judgments of Auditors

Judgment is an inferential cognitive process (Rohrbough 1979) and it permeates every audit task and engagement. Auditors make judgments and decisions about unknown quantities (such as some account balances) or qualities (such as the risk of misstatement of an account balance, going concern status, and the strength of internal controls) on the basis of available information (cues). Because judgment is extremely common and important for auditors and accountants, particularly in uncertain environments, the investigation of their judgment processes has been widespread (see for example, R. Ashton and A. Ashton 1995; Trotman 1996). There are two major purposes of judgment research in accounting and auditing. First, the research contributes to the knowledge of judgment processes. Through this knowledge accumulation, we have a better understanding of judgment processes. Second, the ultimate objective is to improve accountants' or auditors' judgments (Libby 1981 p.3). There are a number of means to improve auditors' judgment. Training and instruction, the review process, provision of feedback, and use of decision aids are some commonly used methods. Their effectiveness in improving auditors' judgment would depend on a number of task contextual and individual variables. Therefore, research in assessing the effectiveness of different means of improving judgments and decisions, and the conditions under which these means are effective, is important in auditing. One of the important task contextual variables affecting auditors' judgment is task complexity.

There are studies in both the accounting and psychology literatures investigating the effect of various kinds of feedback on judgment performance. Outcome feedback (OFB) is the knowledge of the outcome of a judgment (which may or may not immediately follow the judge's prediction). Task properties feedback (TPF) provides information about the formal properties of the task, such as the optimal weights attached to various cues in order to accurately predict the criterion. Cognitive feedback (CFB) refers to information about an individual's judgment strategy, such as the weights that the individual attached to the various cues.

After reviewing the audit judgment literature, Bonner (1994) concluded that research examining "the effects of audit task complexity on auditors' judgments has been minimal" (p.213). She developed a model on audit task complexity to facilitate future research, and interpret and generalise research results. Study of task complexity is important for several reasons. First, it is important to understand audit task complexity and its effects on judgment performance. Even seemingly minor changes in trivial task characteristics can affect judgment (Libby 1981). A crucial task characteristic like complexity can have material impacts on audit judgments, especially because audit tasks vary largely in complexity. Second, understanding audit task complexity is useful for us to interpret research results as well as auditors' performance in audit judgment tasks. Third, based on the level of task complexity, appropriate decision aids and training can be devised. In terms of audit planning, knowledge of task complexity is crucial in job assignments (Abdolmohammadi and A.Wright 1987; Libby 1981) and in performance evaluation. Task complexity also affects the effectiveness of different types of feedback and feedback contents. Steinmann (1976) notes that cognitive feedback may be more useful than task properties feedback when tasks are complex. Balzer et al. (1992) urge more research into the effect of cognitive feedback and outcome feedback on judgment performance given the mixed results across studies, especially with more realistic substantive tasks. In addition, Balzer et al. (1994) suggest that research should be conducted such that "...experienced, professional judges be recruited to participate in future studies" (p.383).

According to Bonner's (1994) audit task complexity model, three components of information processing stages, i.e., input, processing, and output, would affect task complexity. Within each of these stages, either the amount of information or clarity of information is relevant to task complexity. In this study, one aspect of task complexity in Bonner's (1994) model - the clarity of information in the processing stage – was manipulated. Specifically, the functional form of information processing was manipulated at two levels. The low complexity task in this study involved linear processing of five cues, whilst the high complexity task required configural cue processing of two of the five cues. Most previous feedback research in accounting and auditing required subjects to process information with linear functions (e.g. Harrell 1977; Hirst and Luckett 1992; Kessler and R. Ashton 1981). However, a lot of accounting and audit judgment tasks require configural cue processing, such as internal control strength and risk of misstatement of account balance judgments (Brown and Solomon 1990; 1991). I assessed the effects of different types of feedback on auditors' judgment performance under two levels of task complexity.

1.2 Aims of the Research

The main aim of this study is to investigate the effect of different types of feedback (outcome, task properties, cognitive, and combined task properties and cognitive), under different levels of task complexity on audit judgments. It manipulates an element of task complexity in Bonner's (1994) model - the clarity of processing (operationalised as the functional forms of cues). In addition, the study investigates the effect of different types of feedback on the transfer of learning. The purpose is to assess whether the learning from different feedback types in one audit task can be transferred to another audit task. This is an issue that has not been studied in the accounting and auditing literatures. Further, this study also investigates the moderating effect of self-insight of different types of feedback in improving judgment. This issue has not been addressed in the psychology, accounting, or auditing literature.

This study extends previous feedback research in four main aspects:

1. Different types of feedback are examined across two levels of task complexity. The types of feedback examined are: no feedback, outcome feedback, task properties feedback, cognitive feedback, and combined task properties and cognitive feedback. Task complexity is explicitly manipulated in the experiment and the manipulation is guided by Bonner's (1994) audit task complexity model and suggested by Steinmann (1976) and Balzer et al. (1992).

Task complexity is manipulated by varying the functional form of information processing. Specifically, high task complexity involves configural information processing and low task complexity involves linear information processing.

2. While outcome and task properties feedback have been studied in the accounting / auditing literature, cognitive feedback and its combination with task properties feedback have not.
3. This study extends the feedback research by considering the transfer of learning. That is, it investigates whether the learning from various types of feedback can be transferred to subsequent audit judgment tasks. It is generally agreed that the learning and training effect is highest when the learning can be transferred to tasks of both similar and different natures (Haccoun 1997).
4. This study extends the feedback research by investigating the effect of self-insight in moderating the effectiveness of different types of feedback in improving judgment in tasks of different levels of complexity.

This study has two subsidiary aims. First, this study examines the effect of various feedback conditions on promoting the extent of configural cue processing if the domain- and task-specific knowledge prescribed that such information processing is appropriate. Second, this study assesses the effect of various feedback conditions in improving the lens model parameters, such as judgment consistency (Rs), knowledge of the linear component (G) and nonlinear component (C) of the task. Analysis of these lens model parameters can identify the sources of improvement in judgment performance under each feedback condition.

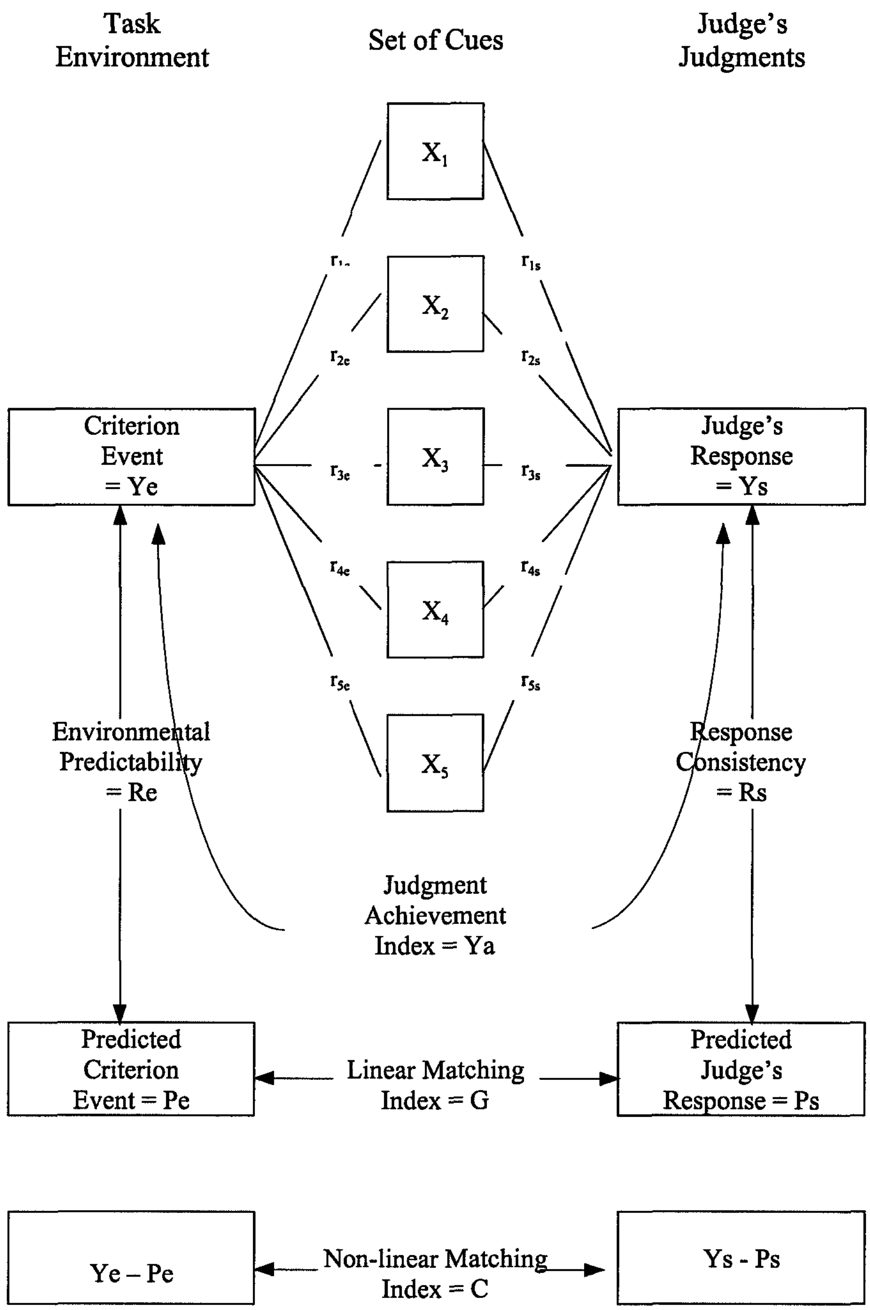
1.3 Theoretical Framework

The major stated aim of this thesis is to evaluate the effectiveness of different types of feedback in improving auditors' judgment performance in audit tasks across two levels of complexity. It is hypothesised that the judgment performance of auditors, such as accuracy, consistency and self-insight, would be different for different types of feedback if the tasks differ in the level of task complexity. The differential effects of different types of feedback on auditors' judgment performance can be shown by utilising the Brunswik Lens Model as the theoretical framework.

Brunswik's Lens model highlights many essential features of judgment under uncertainty (Libby 1981, p.18). From Brunswik's point of view, the environment in which an individual is embedded and the individual himself/herself both affect his/her judgment performance. The model takes the perspective that there is inherent uncertainty in the task environment in that the proximal cues are only probabilistically related to a particular distal state of affairs or events. Also, there is inherent uncertainty within an individual (judge) as how the cues would be utilised to generate functional responses (Cooksey 1996, pp. 2-3).

The Lens model has been widely used to examine judgments about the environment, based on a set of information cues (or pieces of information), which are probabilistically related to a particular event (Trotman 1996, p.33). Diagrammatically, Brunswik's Lens Model can be shown as in Figure 1.1.

Figure 1.1
Simplified Lens Model Framework with Superimposed Statistical Parameters



Judgments about environmental events are usually based on a set of relevant cues. These judgment processes can be modeled in terms of the lens model paradigm. In this model, judgment performance, such as accuracy or consensus, is a function of task knowledge, consistency (cognitive control) and task predictability. An individual's judgment behaviour can be modeled by means of Brunswik's lens model (Hammond et al. 1964; Tucker 1964):

$$r_a = R_e R_s G + C [(1 - R_e^2)(1 - R_s^2)]^{1/2}$$

r_a = judgment achievement / agreement

R_e = environmental / task predictability

R_s = response consistency (cognitive control)

G = linear matching index (knowledge of the linear component of the task)

C = non-linear matching index (knowledge of the non-linear component of the task)

Applying the lens model to an ordinary judgment/decision making context, it is rare for individuals to have access to the true (largely unobservable and unknown) environmental events/states. Instead, the task environment normally provides individuals with a number of information cues upon which they must base their judgments. These information cues are of varying reliability and validity, and they may relate to the environment events in both linear and curvilinear functional forms (including configural patterns). This uncertain task environment gives rise to the zone of ambiguity (in the context and terminology of the lens model), which lies between the observable proximal set of information cues and the unobservable environmental events. It is this zone of ambiguity that leads to different judgment processes and makes judgment tasks more or less difficult.

The zone of ambiguity can be described in terms of at least the following seven parameters:

i. the number of information cues;
ii. the presentation order or format of information cues;
iii. the ecological validity (i.e., weight or importance) of each cue;
iv. the functional form (e.g., linear or curvilinear) between each cue and the environmental event;
v. the organising principle (e.g., additive or configural) of the task;
vi. the predictability of the task; and
vii. the interdependency of cues (i.e., the extent to which the cues are intercorrelated).

It is variation in these properties of the zone of ambiguity in judgment tasks that makes the generalisation of results among studies difficult. Also, it contributes to the precariousness of applying research results to applied settings, such as auditing. For example, it would require a thorough understanding of the domain- and task-specific knowledge and judgment practices of auditors in order to fully describe the parameters of an audit judgment task in the real world. The complexity of an audit task is impacted by its parameters of the zone of ambiguity.

In the present study parameter iv, the functional form between cues and the criterion, is manipulated. The two levels of this variable are configural and non-configural relationships between cues and the criterion.

Figure 1.1 can be conceptualised as lens model's representation of a judgment situation where an auditor determines the probability (or risk) of a material misstatement of the accounts receivable balance in the financial statements, based on the outcome of five audit procedures (cues). The outcome of these audit procedures is, however, an imperfect indicator of the probability of a material misstatement of the accounts receivable balance in the financial statements. In effect, the auditor observes and predicts the environment or events through a "lens" of imperfect cues. The judgment performance / achievement is shown as a function of the environment, the task, and the judge (auditor). In order to understand potential improvements in judgments, we should study all of these three elements jointly.

Like many judgment tasks, judging the risk of misstatement of the accounts receivable balance may not have an objective criterion. In this case, some studies have used the consensus response of a panel of experts as a criterion (for example, Libby and Libby 1989). This approach has been adopted in this thesis.

Judgment achievement (r_a) is the correlation between an individual's judgments and the environment (external criterion). If it is the correlation with another person's judgment, it is called agreement. The lens model equation indicates, that in order for an individual to be able to predict some external criterion: that external criterion (Re) itself must be predictable; he/she must apply his/her judgment policy in a consistent manner (R_S); and he/she must have accurate knowledge of the linear (G) and non-linear (C) components of the environment. A judge can improve his/her judgment accuracy by improving his/her knowledge of the task (both G and C), and/or applying the knowledge more consistently. In addition to judge-related variables, his/her judgment accuracy also depends on the task predictability.

Within the Brunswik Lens Model, policy capturing research in auditing has been one of the dominant paradigms (Trotman 1996, p.41) and it has been used in hundreds of studies of judgment and decision making in the psychology literature (Steward 1988). The main objective of this type of research is to develop mathematical representation of judgment policies, to describe quantitatively the relationship between judgments and information (cues) used to make that judgment. This thesis also adopts the policy capturing paradigm because one of its objectives is to evaluate the effectiveness of different types of feedback in improving auditors' judgment performance in terms of accuracy, consensus, cue usage and the use of specific patterns of cue combination in their judgment policies.

1.4 Auditing Tasks

The two auditing tasks used in this study are the judgment of the risk of misstatement of accounts receivable in the financial statements and the judgment of the control risk of a cash disbursements system. The details of the two auditing tasks will be outlined in the Research Methods section. The accounts receivable task is adapted from Brown and Solomon (1991, Experiment 1) and the cash disbursement task is adapted from Brown and Solomon (1990). The tasks have been carefully developed, tested, and used in prior studies (Brown and Solomon 1990; 1991; Hooper and Trotman 1996).

1.5 Structure of the Thesis

Chapter two contains the literature review which considers the major areas of literature relevant for hypotheses development. Chapter three presents the development of hypotheses addressing the impact of different types of feedback on judgment performance for two audit tasks of different levels of task complexity, the moderating effect of self-insight, and the transfer of learning.

Chapter four outlines the research design employed by the study and describes the experimental tasks. The definitions and measurement of variables are discussed. This chapter also contains the contents and presentation of different types of feedback. Chapter five presents the results of analyses relating to the hypotheses. The results of additional analyses, such as the effect of different types of feedback on promoting configural cue processing and Lens Model parameters, are also presented. Chapter six presents the discussion on results.

Chapter seven summarises the main findings arising from the experiment, discusses the practical and theoretical contributions of the results, outlines some of the limitations to the study, and presents some potential areas for future research.

Finally, two appendices are included in the thesis. Appendix one presents the computer screens of the instructions to and case information of the first block of trials. Appendix two shows the presentation of outcome feedback.

CHAPTER 2

LITERATURE REVIEW

Previous studies relating to feedback and complexity will be reviewed. These represent the key elements in this study. This chapter is divided into two sub-sections: overview of feedback studies and overview of task complexity.

2.1 Feedback - Overview

Feedback is an important construct in a number of disciplines, such as cybernetics, engineering, economics, organisational theory, psychology, and human judgment and decision making. It is generally conceived as the process by which the environment returns to individuals a portion of the information in their response output necessary to compare their present strategy with a representation of an ideal strategy (Doherty and Balzer 1988; Balzer et al 1989). Feedback is thought to improve individuals' judgments by enhancing their understanding of the judgment strategies, further develop their expertise in the task, and correct the use of inappropriate judgment strategies (Hogarth 1987, p.127).

Judgment permeates auditors' work. Auditors have to make a series of judgments in planning, executing, and reporting in an audit engagement. Therefore, auditors have to consistently apply their domain- and task-specific knowledge and an appropriate judgment strategy in order to achieve superior judgment performance. The psychology and accounting literatures show that feedback provides one means for learning the formal properties of the task and the relationship between cues and outcomes. As such, feedback aids in developing an appropriate judgment strategy. However, the literature shows that the outcomes vary considerably with the type of feedback and other environmental factors.

There are studies both in the accounting and psychology literatures investigating the effect of various kinds of feedback on judgment performance. As noted in Chapter 1, outcome feedback (OFB) is the knowledge of the outcome of a judgment. Task properties feedback (TPF) provides information about the formal properties of the task, such as the optimal weights attached to various cues required in predicting the criterion. Cognitive feedback (CFB) refers to information about an individual's judgment strategy, such as the weights that the individual attached to the various cues. Studies in the accounting literature have focused on a number of accounting-related judgment/decision tasks, for example, budgeting (Cook 1968), staff performance evaluation (Harrell 1977; Hirst and Luckett 1992), product pricing (R. Ashton 1981), bond-rating (Kessler and R. Ashton 1981; R. Ashton 1990), and bankruptcy prediction (Hirst, Luckett and Trotman 1999). Although the tasks are quite diverse in these studies, they have several common features (Kessler and R. Ashton 1981). First, the judge or decision-maker uses accounting numbers (or cues) to predict or judge the value of a current or future event (or criterion). Second, the judge / decision-maker has to integrate several pieces of information (or cues) in making the judgment. Third, each cue is probabilistically related to the variable of interest (a current or future event). Fourth, the cues have different predictability. Fifth, the probabilistic nature of the task and environment prevent perfect prediction.

The main focus of most studies has been on the relative effectiveness of different types of feedback (and combinations of feedback) in improving judgment performance. Judgment performance is related to changes in behavioural criteria, such as knowledge, cognitive control, and achievement (or accuracy). These criteria are not independent. In most judgment tasks, achievement is the most important criterion. Achievement is the accuracy of predicting the environment of interest or the consensus with another person or certain benchmark.

The types of feedback examined in both the accounting and psychology literatures mainly focus on the effect of outcome and task properties feedback on judgment performance. There are some studies in the psychology literature investigating the effect of cognitive feedback on: judgment accuracy and validity in a multiple-cue probability-learning task (MCPL) (Balzer et al. 1994); production planning (Remus et al. 1998); forecasting time series data (Remus et al. 1996); interpersonal learning in negotiation (Thompson and DeHarport 1994); and group performance (Harmon and Rohrbough 1990). Only one accounting-related research study investigated cognitive feedback. Kessler and R. Ashton (1981) studied the effect of TPF and CFB in a bond-rating task with MBA students as subjects. At this point, I am not aware of any study investigating and comparing the effectiveness of CFB in the audit context.

Below I outline the main psychology and accounting studies that have addressed the types of feedback considered in this thesis.

2.1.1 Outcome Feedback (OFB)

OFB tells individuals whether their judgments lead to success or failure (or some variation thereof) but nothing about the reason for their occurrence. With a few exceptions (e.g., Adelman 1981; Schmitt et al. 1976; Schmitt et al. 1977), OFB is not effective in improving judgment performance in the psychology literature (Balzer, Doherty and O'Connor 1989; Einhorn and Hogarth 1981), especially in complex inference tasks. For example, in 1968, Goldberg (1968) reported that:

> "...there was virtually no cross-sample generalization of learning as a result of intensive training on over 4000 MMPI profiles..." (p.492)

More recently a similar conclusion has been made by Balzer et al. (1989) and Hammond (1996, p. 270).

The usefulness of OFB in improving learning under uncertainty is doubtful (Hammond 1996, p. 270) and it is suggested that OFB may hinder learning (Brehmer 1980; Hammond et al. 1973). If the task and/or the OFB has random errors, an individual may not be able to develop an appropriate strategy from the outcomes or may infer incorrect judgment strategies due to noise or random errors in the outcomes (Bonner and Walker 1994). That is, the decision-maker loses cognitive control and he/she would find it difficult to distinguish between systematic and random cue-criterion associations (Hammond 1971). Also, OFB is generally ineffective because it provides no direct information on the nature or extent of correlation between each cue and the criterion. The task environment together with limited human information processing capability makes learning (i.e. the development of an appropriate judgment strategy to improve quality of judgment) difficult. Further, the inability to gather OFB in certain circumstances (such as on rejected alternatives) would also impede its effectiveness in improving judgment.

However, some studies found that OFB could be effective in improving judgment accuracy and consensus in tasks, such as performance evaluation in accounting firms (Hirst and Luckett 1992; Luckett and Hirst 1989). It should be noted that OFB is useful in improving judgment only under certain conditions where: (i) the task is of high predictability; (ii) the task has a small number of meaningfully labeled cues; and (iii) judges have high initial domain-specific knowledge and causal theory about the task. The first condition is that the task needs to be relatively simple and have high task predictability so that individuals can reason backwards from the OFB and develop the appropriate judgment strategies (Hirst and Luckett 1992). Simple tasks usually are of high predictability, involve a smaller number of cues (Bonner and Walker 1994; Hirst and Luckett 1992), and / or require only a few stages to be undertaken to complete (Wood 1986). Goodman (1998) found that OFB is beneficial to performance in a cognitive problem solving task with 202 undergraduate subjects. Schmitt et al. (1977) found that OFB is useful in improving college students' matching, but the difference is not large enough to offset the lack of consistency in a student achievement prediction task using three correlated cues across two blocks of 25 profiles. There are studies in the accounting literature that support this notion. Airforce officers learned from

OFB in a performance evaluation task that had high environmental predictability, and where there were no random errors and noise in the OFB (Harrell 1977). Hoskin (1983) found that OFB improves forecasting product demand. In a performance evaluation task, Hirst and Luckett (1992) found that auditors learn from OFB and their performance improves over time.

The second condition is that the task has a small number of meaningfully labeled cues. In a bond rating task involving only a linear combination of three accounting ratios, auditors learned from OFB (R. Ashton 1990).

The third condition for OFB to improve judgment performance is that judges must have high initial domain- and task-specific knowledge (Hirst and Luckett 1992) and a causal theory about the task (Bonner and Walker 1994). This prior knowledge and causal theory enables judges to analyse and interpret OFB, and assess the correctness and appropriateness of their theories and judgment strategies (Carmerer 1981). In studies where subjects had high domain- and task-specific knowledge about the task, it was found that OFB facilitated learning and improved judgment performance (Adelman 1981; Hirst and Luckett 1992; Hirst et al. 1999). Bonner and Pennington (1991) also found that auditors' performance was affected by the amount of their prior instruction, which was related to their prior knowledge.

Given these three conditions, individuals can assess the correctness and appropriateness of their judgments and judgment models from the OFB. This information enables them to detect incorrect or inappropriate judgment strategies. Hirst and Luckett (1992) found that OFB does improve performance evaluation judgment over time. Auditors receiving OFB alone, and in combination with TPF improve judgment accuracy for both high and medium levels of task predictability (Hirst et al. 1999). In an extensive review of psychology literature, Kopelman (1984) found that OFB can enhance performance in many motor skills and simple tasks.

In the auditing context, Bonner and Pennington (1991) concluded that OFB was not effective in improving auditors' judgment after examining training schedules, seeking professional guidance, and reviewing materials used in large audit firms, as well as surveying audit partners and managers concerning the knowledge related to different tasks, learned from experience or training. Most of the auditing tasks do not satisfy the above-mentioned three conditions and therefore OFB is unlikely to be effective.

2.1.2 Task Properties Feedback (TPF)

Task properties feedback refers to the relationship between the cues and the criterion, information about the criterion, and /or the cues themselves (Balzer et al. 1989). In short, TPF is statistical information about the task. The strength of association between each cue and the criterion is an example of TPF information. TPF gives a guideline with regard to the optimal judgment strategy and the manner in which processing strategies might advantageously be changed.

Earlier psychological studies demonstrate that TPF alone is effective in improving judgment performance (Newton 1965, Schmitt et al. 1976), when compared with no other information. Deane (1970, cited in Hammond and Boyle 1971) conducted an experiment with 50 college students with an abstract task using three curvilinear cues and found that TPF presented in graphical forms led to an improvement in performance over OFB. Nystedt and Magnusson (1973) found TPF effective in improving the accuracy of predictions of achievement scores, on the basis of the results of five tests (differed in ecological validities), when compared with the no other information group. Schmitt, Coyle and King (1976, Experiment 2) found TPF improved college student's judgment performance in predicting student achievement and Remus et al. (1996) found TPF improved performance in a dynamic forecasting task. Research has indicated that learning and improvement in judgment performance are more effective if subjects are shown which cues are valid and to what extend they are valid. Individuals can quickly utilise that information in making their judgments (Hammond 1996, p. 270).

Balzer et al. (1992) found that TPF alone, and with other types of feedback, improved judgment accuracy and validity in a MCPL task. Also, they found no further improvement in judgment by adding other types of feedback, such as cognitive feedback. Based on their results, they concluded that TPF was the major factor contributing to judgment performance improvement. Remus et al. (1996) found similar results in a task of forecasting time series data with fundamental changes and built-in randomness.

R. Ashton (1981) studied OFB and a form of TPF and found that neither type of feedback had a significant effect on the acquisition or application of decision model knowledge in an accounting task involving making pricing decisions. R. Ashton (1981) suggested that one possible reason was a ceiling effect because the pre-feedback performance level was so high that there was little room for improvement.

Harrell (1977) demonstrated that providing senior military officers with a TPF "official" policy statement indicating the relative importance of the five goals in pilot training in performance evaluation of trainee pilots, improved their matching of importance weights for two of the five evaluation criteria.

Kessler and R. Ashton (1981) concluded that in a corporate bond-rating task with three accounting ratios involving 69 MBA students, TPF was more effective than CFB in promoting consistency in judgment and it was as effective as CFB in improving accuracy and knowledge acquisition. Given that providing TPF is relatively cheaper than providing CFB, the former was suggested as the preferred feedback form.

In a performance evaluation task involving practising auditors, Hirst and Luckett (1992) showed that TPF promoted judgment achievement. In a task involving the prediction of corporate failure using audit seniors from Big-6 accounting firms, Hirst et al's. (1999) results showed that TPF did not promote judgment accuracy above that of the control group when task predictability was high. When task predictability was at a medium level, TPF led to higher judgment accuracy than that of the control group.

Bonner and Walker's (1994) results supported the hypothesis that practice with explanatory feedback (a form of TPF) promotes procedural knowledge acquisition irrespective of the form of instruction received by subjects (95 senior undergraduate and graduate students) in an auditing task involving predicting error based on four financial ratios. However, their results also indicated that subjects receiving and understanding rules and declarative knowledge instructions prior to practice with OFB can also promote procedural knowledge acquisition which is comparable to subjects receiving explanatory feedback.

2.1.3 Cognitive Feedback (CFB)

Research results on the effectiveness of cognitive feedback are not as clear as the results of TPF. Galbraith (1984) found significant improvement in cognitive control when subjects received CFB rather than TPF. Some studies in psychology found that CFB and TPF are equally effective (e.g., Steinmann 1976). Schmitt et al. (1976) conducted two experiments that manipulated three feedback conditions (OFB, TPF, and CFB) and examined their impact on judgment accuracy, matching and consistency. In two experiments (with tasks varying in predictability), there was no significant CFB main effect on these three feedback conditions. However, in the high predictability task (Experiment 1), the presence of CFB together with TPF improved judgment accuracy and matching significantly. They concluded that

CFB, when supplied with TPF, was effective in improving judgment in situations where the task predictability was high and the cues had substantially different importance (cue weights). More recent psychology studies tend to conclude that CFB is not as effective as TPF (Balzer et al. 1992; Balzer et al. 1994; Harmon and Rohrbaugh 1990).

Kessler and R. Ashton (1981) in a bond rating task found CFB equalled TPF in prediction accuracy and judgment strategy but was not as good as TPF in terms of judgment consistency. On the other hand, CFB was found to be useful in task learning (Hammond 1987; Hammond et al. 1977) and more effective in promoting judgment performance (Schmitt et al. 1976).

The effectiveness of CFB in improving judgment performance is an unsettled issue. There has been some speculation about the uses of CFB. After reviewing the feedback literature, Kessler and R. Ashton (1981) speculated that "... certain types or combinations of feedback are more effective for some kind of tasks than for others" (p.148). Newton (1965) asserted that CFB might be beneficial to sophisticated clinicians. Galbraith (1984), Kessler and R. Ashton (1981) and Steinmann (1976) speculated that CFB did not lead to improvement in judgment performance over and above that of TPF, which might be because of the ceiling effect. That is, TPF achieved such a significant improvement, that it left little potential for CFB to demonstrate further improvement. Steinmann (1976) noted that cognitive feedback might be more useful than task properties feedback when tasks were more complex, such as where there is an increase in the number of cues, cue intercorrelations and functional forms. CFB was originally developed partly because evidence suggested that individuals lacked insight into their own judgment policies, that is, a low correlation between self-reported subjective policy weights and statistical weights. This was interpreted as poor understanding or implementation of the judgment policy. Also, subjects' descriptions of their policies were often inaccurate (Balke et al. 1973; D. Summers et al. 1970). Generally, they overestimated the number of cues that were important to them (Slovic and Lichtenstein 1971) and underestimated the weight of those cues that were important to them (Slovic 1969). These findings suggest the possible usefulness of CFB in improving judgment performance, especially in highly complex tasks.

2.1.4 Combined Task Properties and Cognitive Feedback (CombFB)

There are some studies which have investigated the joint effect of TPF and CFB (combined feedback) on improving judgment performance. Balke et al. (1973) found that CombFB improved judgment accuracy when compared with a no feedback condition. Adelman (1981, Experiment 2), found that CombFB resulted in greater knowledge acquisition and judgment accuracy than OFB in a task predicting GPA of college students using four personality variables. Fero's (1975) clinical practitioner subjects had favourable reactions to combined feedback and their knowledge acquisition in the task was significantly better than in the no feedback condition in predicting the drug use and employment status of patients. CombFB, when compared with OFB, also led to improvement in task knowledge in other judgment tasks, such as: in an abstract task using three curvilinear cues

(Hoffman et al 1981); in eight different tasks using three cues with college students as subjects (Steinmann 1974); and in a prediction of student achievement task (Nystedt and Magnusson 1973). When compared with OFB, CombFB generally leads to superior judgment accuracy (Hammond and Boyle 1971; Hoffman et al. 1981; Lindell 1976; Stang 1985; Poses et al. 1986). In respect of cognitive control, CombFB was found to be significantly superior to OFB (Adelman 1981, Experiments 1 and 2; Balke et al. 1973; Fero 1975; Galbraith 1984; Steinmann 1974). However, there are studies indicating that combined TPF and CFB are not better than OFB in promoting cognitive control (Clover 1979; Schmitt et al. 1976, 1977) in college students performing a medical diagnosis task using 2 cues.

However, CombFB normally does not improve judgment accuracy over and above that achieved by TPF (Galbraith 1984; Steinmann 1974; Balzer et al. 1992; 1994). In a bond rating task with three cues, Kessler and R. Ashton's (1981) MBA student subjects found TPF is as effective as CombFB for improving prediction accuracy and task knowledge acquisition.

The conclusion of previous psychology studies on the relative effectiveness of CombFB should be considered with care and as tentative for several reasons. First, the conclusions are based on only a small number of studies that provide a comparison of different components of TPF and CFB. Second, the studies do not provide feedback on the same components of TPF and CFB for subjects and this makes comparison between studies difficult. Third, the sample size in most previous studies has been small (for example, five of the seven studies, which compare the relative effectiveness of TPF and CFB, reviewed by Balzer et al. (1989) had 25 or fewer subjects). Therefore, Balzer et al. (1992) urged more research into the effect of CFB and OFB on performance, given the mixed results across studies.

2.1.5 Feedback Studies with Tasks Involving Configural Cues

There are not many feedback studies with tasks involving configural cues. In a study not involving feedback, Brehmer (1969) found that subjects were slower in learning the configural relationship of cues, but subjects achieved the same level of performance in both linear and configural tasks over 400 trials. Brehmer (1969) speculated that linear relationships were more common in reality and therefore led to different rates of learning for tasks with linear cues as opposed to configural cues. Dean et al. (1972, Experiment 1) had similar results in an experiment with 60 undergraduates and found that judgment performance was higher for linear tasks than for non-linear tasks after receiving OFB. Castellan and Swaine (1977) also found subjects took longer to learn a task involving configural cue processing. There was no difference in performance between OFB, and combination of OFB and TPF conditions. D. Summers and Hammond (1966) found a similar result that non-linear combination of cues was more difficult to learn. Also, they found that only detailed instructions provided to subjects (similar to TPF) would promote the use of appropriate non-linear cue weights.

Hammond and D. Summers (1965) demonstrated that subjects were able to solve problems of inductive inference after completing 100 trials and receiving OFB in either a linear or configural (patterned) manner for a task involving two cues. Further, they found that when subjects were told of the functional form to combine the two cues, they increased their utilisation of configural cue processing. S. Summers, R. Summers and Karkau (1969) also found that undergraduate students could learn the use of configural cue processing after extensive trials with OFB in a task involving two cues.

Newton (1965) reported the results of an experiment where 99 students estimated the admission status of applicants based on five inter-correlated cues under various feedback conditions (OFB, TPF, CFB, and CombFB). The nature and extent of the configuration of the cues was not reported. Newton (1965) found that subjects became more linear in their judgment models in each feedback session, regardless of the type of feedback. Further, subject's use of configural cue processing did not vary systematically with different types of feedback although some subjects relied on configural cue processing to an important extent.

In the psychology literature, one of the overwhelming findings is that individuals tend to process information or cues in a linear manner (e.g., Goldberg 1968; Slovic et al. 1977). Brown and Solomon (1991) drew the following conclusion after reviewing the relevant literature:

> "With few exceptions, prior policy-capturing psychology studies have *not* produced evidence that individuals, either experts or novices, configurally process available information..." (p. 102).

Also, prior policy-capturing studies in accounting and auditing essentially found no evidence that accountants or auditors configurally processed information or cues (R. Ashton 1974; R. Ashton and Kramer 1980; Hamilton and W.Wright 1982; Hall et al. 1982; Trotman et al. 1983). One of the reasons for the inability to detect configural information processing could be that the tasks employed by these studies did not require such an information processing mode. Brown and Solomon (1991) reviewed the procedures and materials employed in prior auditing studies and noted that

> "... it was observed that none of the studies captured two or more control features that related to the same control objective in such a way that configural information processing would be a salient, appropriate strategy. Consequently, there is little reason for an auditor to have done anything other than what was observed in prior studies: evaluate independently the effect of each control" (p.103).

Brown and Solomon (1990, 1991) designed auditing tasks where domain- and task-specific knowledge required configural information processing to make a judgment and found that a significant number of auditors did employ configural information processing. They concluded that auditors' information processing was more complex than has been reported by prior studies. I am not aware of any auditing research investigating the effect of various feedback conditions in improving the extent of auditors' configural information processing.

2.1.6 Feedback and Self-insight

An overwhelming finding in the psychology literature is that individuals often describe their judgment policies inaccurately (Balke et al. 1973; Reilly and Doherty 1992; D. Summers et al. 1970). Reilly and Doherty's (1992) comment summarised the results of this line of research: "A widely accepted finding in research on human judgment is that people have relatively poor insight into the weighting schemes they use when they make holistic judgments" (p.285). Individuals overestimate the number of cues they used in their judgment models (McKenzie 1997; Slovic and Lichtenstein 1971). They also overestimate the weights of those cues that are not important to their judgments (Slovic 1969).

A few studies using accounting related tasks also found rather low self-insight for professional subjects, such as security analysts (Mear and Firth 1987), and stockbrokers (Slovic Fleissner and Bauman 1972). Colbert (1988) found that auditors had a fair amount of self-insight in assessing inherent risk using four cues. Jiambalvo, Watson and Baumler (1983) also had a similar result in a performance evaluation task with 152 senior auditors and staff accountants. They noted that "individuals appeared to possess a fair amount of insight into their evaluation policies, but, in general, they overestimated the importance of minor evaluation categories and underestimated the importance of a few major categories" (p. 24). However, in R.Ashton's (1974) internal control evaluation task involving six cues, the auditor subjects exhibited high self-insight. In replications and extensions of R. Ashton's (1974) study, R. Ashton and Kramer (1980), R. Ashton and Brown (1980), and Hamilton and W.Wright (1982) also found that subjects had high self-insight. One of the reason for such a high level of self-insight may be due to the fact that evaluation of internal control systems is a well-structured judgment task (W.Wright 1988).

A few accounting studies investigated the effect of OFB and TPF on self-insight. The effect of CFB and CombFB on self-insight has not been studied. Luckett and Hirst (1989) found that auditors receiving either OFB, TPF, or both OFB and TPF on a performance evaluation task improved their self-insight. However, Tuttle and Stocks (1997) had quite different results in a financial distress prediction task based on six orthogonal financial ratios with 71 accounting students. They found that in the absence of OFB or TPF, the student subjects showed significant insight. When TPF or OFB was provided, subjects' self-insight disappeared. They attributed the results to the fact that subjects confused the TPF and OFB with their own decision models, which is in line with the explanation proposed by Nisbett and Wilson (1977). However, previous research has not considered the relationship between self-insight and the impact of cognitive feedback. This is an important issue given that CFB was partly developed as a means to improve individuals' self-insight.

Table 2.1 summarises the accounting and auditing studies that have addressed feedback and some of the issues investigated in this thesis.

Table 2.1
Summary of Feedback Studies in Accounting

Studies	Types of Feedback Investigated	Task	Subjects	Is Control Group Employed?	Is Task Complexity Manipulated?	Is Task Involving Configural Cues?	Is Self-insight Investigated?	Is Impact on Future Judgment Investigated?
Harrell (1977)	NFB TPF TPF + OFB	Performance evaluation of pilot training wings (5 cues)	75 Military Officers	Yes	No	No	No	No
Ashton (1981)	TPF "General" FB (textbook descriptions of the 3 variables)	Setting selling price (3 cues)	36 accounting students; 56 MBA students; and 46 doctoral students	No	Task predictability is manipulated at 3 levels	No	No	No
Kessler & Ashton (1981)	Performance Feedback (no. & % correct and no. & % off by 1 rating) TPF + Performance FB CFB + Performance FB TPF + CFB + Performance FB	Bond rating (3 cues)	69 MBA students	No	No	No	No	No

Table 2.1 (cont'd)
Summary of Feedback Studies in Accounting

Studies	Types of Feedback Investigated	Task	Subjects	Is Control Group Employed?	Is Task Complexity Manipulated?	Is Task Involving Configural Cues?	Is Self-insight Investigated?	Is Impact on Future Judgment Investigated?
Hoskin (1983)	OFB	Inventory ordering task (3 cues)	61 MBA/PhD students	No	No	No	No	No
Luckett & Hirst (1989)	OFB TPF OFB + TPF	Performance evaluation of audit assistants (5 cues)	48 Big-8 auditors	Yes	No	No	Yes (allocate 100 points)	No
Ashton (1990)	OFB	Bond rating (3 cues)	182 auditors	Yes	Task difficult is effected by making available a decision aid that implies a high performance standard	No	No	No

Table 2.1 (cont'd)
Summary of Feedback Studies in Accounting

Studies	Types of Feedback Investigated	Task	Subjects	Is Control Group Employed?	Is Task Complexity Manipulated?	Is Task Involving Configural Cues?	Is Self-insight Investigated?	Is Impact on Future Judgment Investigated?
Hirst & Luckett (1992)	NFB OFB TPF OFB + TPF	Performance evaluation of auditors	48 auditors	Yes	No	No	No	No
Bonner & Walker (1994)	OFB OFB + Explanatory FB (Explanatory FB is similar to TPF)	Identify errors in analytical procedures (4 cues)	95 undergraduate and graduate students	No	No	Yes, but not specifically stated	No	No
Tuttle & Stocks (1997)	OFB TPF (TPF: rank of importance of cues only)	Bankruptcy prediction (6 cues)	71 accounting students	No	No	No	Yes (allocate 100 points)	No
Tuttle & Stocks (1998)	NFB OFB TPF OFB + TPF	Bankruptcy prediction (6 cues)	71 accounting students	Yes	No	No	Yes (allocate 100 points)	No

Table 2.1 (cont'd)
Summary of Feedback Studies in Accounting

Studies	Types of Feedback Investigated	Task	Subjects	Is Control Group Employed?	Is Task Complexity Manipulated?	Is Task Involving Configural Cues?	Is Self-insight Investigated?	Is Impact on Future Judgment Investigated?
Hirst, Luckett & Trotman (1999)	NFB OFB TPF OFB + TPF	Corporate failure prediction (3 cues)	136 audit seniors	Yes	Manipulated at two levels of task predictability	No	No	No
Earley (2001)	OFB + Explanatory FB (Explanatory FB is similar to TPF)	Real estate valuation	159 auditors	No	No	No	No	No

2.1.7 Feedback and Transfer of Learning

The goals of training and learning in real-world settings are to increase the level of post-training performance in the long-term and to transfer the knowledge and skills acquired to related tasks and altered contexts (Schmidt and Bjork 1992). The second goal mentioned above refers to the learning from one task enhancing the performance in a somewhat different context / task (Salomon and Perkins 1989). This is generally referred to as transfer of learning or inter-task learning. The importance of the capability to transfer acquired knowledge and cognitive skills to novel tasks and problems is recognised with respect to learning in both universities (Bridges 1998), and in professional training (De Corte 1996).

The extent of the transfer of learning is affected by the transfer distance, which relates to the degree to which the transfer task is remote or novel in comparison with the original learning / training task. In general, the greater the transfer distance, the more difficult is the transfer. Thus, task similarity affects the transfer distance. The transfer of learning to a similar task is considered a "near" transfer, whilst the transfer to a less similar or novel task is considered as a "far" transfer.

Despite the importance of transfer of learning, there have been very few studies investigating the effect of feedback in promoting such processes. Balzer et al. (1989) found only one previous study (Steinmann 1974) that investigated the effect of training and learning with OFB, TPF, and CombFB on judgment tasks on the transfer of learning to other tasks. McCarthy et al. (1995) made a similar comment after reviewing the education psychology literature and noted that "Minimal research related to feedback and transfer has been published..." (p. 142). As concluded by Steinmann (1974):

> "Research pursuing the possibility that policy formation may transfer, under some conditions, seems essential; heterogeneity of judgmental conditions is characteristic of real life situations..." (p.15).

Steinmann (1974), in an experiment with 16 students, found that providing TPF and CombFB led to an increase in the transfer of learning (an increase in accuracy) to subsequent tasks that differed in task predictability and cue ecological validity. McCarthy et al (1995) found that OFB assisted students in acquiring a base for transfer in a near-transfer task.

2.1.8 Shortcomings of Previous Psychology Feedback Research

There are a number of shortcomings of previous feedback research, such as the tasks used, subjects, feedback contents, and feedback presentation formats which are described below. These make generalisation and comparison of results across studies very difficult. Care must be exercised in interpreting the results.

2.1.8.1 Task

It was noted by Balzer et al. (1989), after an extensive review of the feedback research in the psychology literature, that most psychology studies used fairly simple generic "judgment tasks (e.g. a small set of uncorrelated cues that are related to the criterion in an additive fashion)" (p.429). This point was also raised in Balzer et al. (1992) "[T]he failure to find a significant difference among the feedback conditions.... was not completely unexpected.... given that the judgment task included only linear cue-judgment relations. Thus, given *the lack of* *configurality in the judgment task*, there was no reason to expect that feedback would have affected judges' levels of configurality (which were all, in fact, low)" (emphases added, p.48). In the real world, "judgment tasks are dynamic and far more complex than those studied" (Balzer et al. 1989, p.429).

2.1.8.2 Subjects

Most feedback studies in the psychology literature used student subjects (Doherty and Balzer 1988; Balzer et al. 1989). Some accounting feedback studies used student subjects as well (Kessler and R. Ashton 1981). In a recent paper, Balzer et al. (1994) urged that "...experienced, professional judges be recruited to participate in future research studies" (p.383).

2.1.8.3 Feedback Contents

The contents of feedback varied across studies and this makes the comparison of results across studies very difficult. For example, the contents of CFB ranged from a single sentence to very detailed cognitive information. In Remus et al. (1996), CFB was simply stated in a sentence : "You are overreacting to the random variation in the data". On the other hand, in Balzer et al. (1992), CFB had information regarding the subjects' (i) consistency of judgment, (ii) relative weights, (iii) the functional form relating each statistic to the predicted number of criteria, (iv) the average number of predicted criteria, and (v) the range of the criteria within 1 and 2 standard deviations. Therefore, it is not surprising to find quite different results.

2.1.8.4 Feedback Presentation Formats

Different presentation methods were used in both psychology and accounting feedback research. Some psychology studies used graphics (Balzer et al. 1994, Fero 1975, Hammond 1971). A few psychology studies used tables and narrative descriptions (Remus et al. 1996). More recent psychology studies have used a mix of graphics and narrative descriptions (Balzer et al. 1992, Balzer et al. 1994). Most accounting studies used a narrative description to present feedback information to subjects (e.g., Bonner and Walker 1994; Earley 2001; Harrell 1977; R. Ashton 1981; Hirst and Luckett 1992; Hirst et al. 1999; Luckett and Hirst 1989). Some accounting studies used a mix of narrative and tabular formats to present feedback information (e.g., Kessler and R. Ashton 1981; Tuttle and Stocks 1997; 1998). There has not been a common approach to present feedback in previous psychology and accounting studies.

2.2 Overview of Task Complexity

Based on an extensive review of feedback and psychology literatures, Balzer et al (1989) drew the conclusion that task complexity and task contents have not been the subject of most feedback / psychology research. It is generally assumed that the feedback method will have the same effect across tasks. One exception is Adelman (1981), who investigated the effect of task congruence (positive or negative) and its interaction with different types of feedback. A significant main effect for task contents was found. Despite this finding, task complexity and task content have not been explicitly considered in most studies. Probably, one of the reasons was the lack of a comprehensive definition or theory of task complexity and its effects

(Bonner 1994). In the psychology literature, studies in task complexity have suffered from the same problem - lacking a standardised definition of task complexity. As a result, there has been a lack of integration of the evidence of task effects from studies in different areas, and in some cases there have been inconsistencies in results (Wood 1986). He speculated that the inconsistencies in results were probably due, in part, to different manipulations of task complexity.

The body of audit judgment literature that directly investigates the effects of task complexity on auditors' judgments has been minimal (Bonner 1994). In most audit judgment studies, task complexity has not been systematically manipulated. Instead, it has generally been used as a *post hoc* explanation for inconsistent results among studies of auditor judgment, particularly those on the effects of knowledge and skills (Colbert 1989; A.Wright 1988). However, tasks in these studies may have varied in dimensions other than complexity. As a result, the extent of effect of task complexity on the difference in results is unclear (Bonner 1994). Some auditing studies have operationalised task complexity along many dimensions, for example, the structure of the audit task (Abdolmohammandi and A.Wright 1987; Frederick and Libby 1986), task predictability (Simnett and Trotman 1989), and information load (Simnett 1992, Experiment 2).

2.2.1 Theoretical Constructs of Task Complexity

It is necessary to identify the elements or factors affecting the complexity of a task in order to categorise and measure task complexity. There have been a number of models of task complexity proposed. Simon (1960) classified judgments or tasks along a continuum with "programmed" and "non-programmed" at the two ends. "Programmed" tasks are mainly repetitive and structured in nature, whilst "non-programmed" tasks are novel and unstructured. Abdolmohammadi and A.Wright (1987) adopted this categorisation in their empirical study of experience effects and task complexity on audit judgment.

2.2.1.1 Beach and Mitchell's (1978) Model

Beach and Mitchell (1978) proposed a contingency model approach to categorise task complexity. They identified four characteristics of tasks, viz., unfamiliarity, ambiguity, complexity, and instability. Complexity referred to the number of components of the task and the degree to which the task would affect further decisions. This approach to task complexity was useful to match decision strategies to different contributions of task characteristics. Also, task complexity did not depend solely on the characteristics of the task. Instead, it depended on a number of issues: (1) the judge/decision maker; and (2) the interaction of the task and the judge/decision maker, such as the perception of the task by the judge/decision maker.

However, this model was narrow in its definition of complexity (Simnett 1992, p.28) in the sense that task complexity was limited to the number of components of the decision problem. Further, task complexity was influenced by the decision-makers' feeling or perception of some task elements. This made the categorisation of task complexity, on an ex ante basis, impossible.

2.2.1.2 Wood's (1986) Model

Wood (1986) developed a model of task complexity, mainly based on a review of management and psychology literatures and tasks. He suggests that three aspects of a task affect its complexity. These aspects were component complexity, co-ordinative complexity, and dynamic complexity. Table 2.2 summarises the characteristics of these three components. The overall task complexity is a function of these three complexity components.

Variation in task complexity appears to lead to changes in the requirements of an individual, in terms of knowledge, cognitive skill, and effort required to perform the task successfully. On the other hand, the three components of task complexity can be considered independent of each other (Wood 1986). That is, the level of one aspect may change without affecting the levels of the other two aspects. In order to determine total task complexity, the form of the relationship between the aspects of complexity has to be known.

The complexity-performance relationship may be curvilinear. When starting at a low level of complexity, an individual may improve his/her performance with a slight increase in task complexity. The performance of an individual may increase up to a certain complexity level. Beyond this level of complexity, the performance of an individual may decline when the task is at a higher level of complexity. This is because the cognitive demand may be beyond the individual's ability to process the information or cues. This condition is generally known as 'information overload' and it would lead to a lower performance. However, it would be very difficult to determine, especially on an *ex ante* basis, the level of task complexity that induces the highest performance for an individual. Also, the level of task complexity to induce the highest performance is different for different individuals. Further, an individual may have different abilities to handle tasks of different natures or different subject domains. Wood's (1986) model was a generic one and was intended to be applicable to almost all kinds of tasks.

2.2.1.3 Bonner's (1994) Model

After extensive review of accounting, auditing, and psychology literatures, Bonner (1994) developed a model of the effects of audit task complexity. Elements of task complexity were classified into the three components of general information processing stages, i.e., input, processing and output. Within each of these three components, attributes of the task that were relevant to task complexity were grouped as relating to either the amount of information or clarity of information. The amount and clarity of information of this classification scheme, respectively, corresponds closely to the notion of task difficulty and task structure used in previous models.

Table 2.2
Summary of Complexity Components of Wood's (1986) Model

Task Complexity Components	Characteristics
1. Component Complexity	This aspect of task complexity has two elements: (i) the number of acts and stages required to complete a task; and (ii) the number of cues that are required in the processing procedure to complete the task. As the number of acts and cues increases, the knowledge and skill requirements to complete the task will also increase.
2. Co-ordinative Complexity	This component relates to how the cues are being processed and the importance of the cues in affecting the judgment / decision or the completion of a task. The sequencing of cues is an aspect of co-ordinative complexity.
3. Dynamic Complexity	This refers to change in the relationship between the cues and the task outcome mainly due to the dynamic nature of a task. Individuals have to adapt to changes in the cause-effect chain or means-ends hierarchy in completing the task.

There are a number of elements affecting the amount and clarity in the input, processing, and output stages of information processing and decision making. These elements are summarised in Table 2.3. Of particular importance to the present study is the functional form of cues (a task element relating to the clarity of information processing). According to Bonner's (1994) model, people tend to think in terms of simple functions such as linear functions, making them clearer to process than other more complex forms, such as configural (interaction of cues). Previous studies tend to suggest that increasing task complexity (such as the interaction of cues) is negatively related to consistency of cue use and judgment performance (Bonner 1994). Two possible causes are suggested, namely, the use of less appropriate strategies, and the sub-optimal weighting of cues.

Of the three elements of overall task complexity, Bonner (1994) predicted that the first two elements would have a greater impact on task complexity for most auditing tasks. This is because the extent of the variation in output complexity would be quite limited. Most auditing tasks would require only a small amount of output, such as a few judgments or decisions (i.e., limited variation in the amount of output). Bonner's (1994) model is used in this study to guide the design and manipulate the complexity of the audit tasks.

Table 2.3
Task Complexity Elements in the Three Stages of Information Processing and Decision Making

Task Complexity Element	Amount/Clarity	Examples	Related Concepts
1. Input Complexity	Amount of Input	▪ number of cues ▪ number of choices/ alternatives ▪ redundancy among cues	Information load
	Clarity of Input	▪ cues specified (or defined) ▪ cues labeled or not ▪ match between manners in which information cues are presented and they are stored in long-term memory	Demand on cognitive resources
2. Processing Complexity	Amount of Processing	▪ number of cues ▪ number of choices / alternative processing approaches ▪ number of steps or procedures	Information load Demand on cognitive resources
	Clarity of Processing	▪ procedures specified or not ▪ interdependency of procedures ▪ task predictability - magnitude of cue-outcome relations - sign of cues - consistency among cues - functional forms of processing	Demand on cognitive resources Environmental predictability
3. Output Complexity	Amount of Output	▪ number of goals or solutions per alternative (e.g., open-ended or closed-ended tasks)	Information load Demand on cognitive resources
	Clarity of Output	▪ goals or outcomes specified or not ▪ presence or absence of objective criteria to assess the output/outcome	Demand on cognitive resources

2.2.2 Task Complexity in Audit Studies

2.2.2.1 Audit Studies Addressing Task Complexity

Few audit studies have specifically addressed task complexity. Bonner (1994) concluded, after an extensive review of the audit literature that "research which directly examines the effects of audit task complexity on auditors' judgments has been minimal; instead, task complexity has played a subsidiary role in the audit judgment research" (p. 213). While task complexity was not manipulated in most previous audit studies, researchers tended to suggest and speculate that a task complexity difference in their research designs contributed to the inconsistency in research results (Bonner 1994). Researchers also considered task complexity might explain results, which were inconsistent with their hypotheses.

There are a few audit studies, which have manipulated task complexity, for example, Abdolmohammadi and A.Wright (1987), surveyed 105 audit partners and managers and asked them to rate the task complexity of six audit tasks. The questionnaire provided the respondents with a simple definition of task complexity, which followed Simon's (1960) structured and unstructured task complexity continuum. Respondents were asked to evaluate the complexity of the six audit tasks and to indicate their assessments on a continuous scale ranging from zero (indicating a structured task) to 27 (indicating an unstructured task). Based on these results of the survey, in a subsequent experimental study, they selected five audit tasks and tested if there were differences in judgments of these tasks between inexperienced (students and junior supervisors) and experienced subjects (senior auditor or higher). They found a significant experience effect in two unstructured tasks and one semi-structured task. Another semi-structured task (evaluation of internal control) showed a marginal experience effect.

Gaummitz et al. (1982) investigated two audit tasks of different levels of ambiguity (internal control evaluation and audit planning judgment). They posited that the audit planning judgment task was more ambiguous and presumably was of higher complexity. However, they found no experience or task complexity effects in terms of judgment consensus.

Simnett and Trotman (1989) manipulated task complexity by altering the strength of the relationship between the cues (accounting ratios) and the criterion (corporate failure status) in a corporate failure prediction task. Specifically, low task complexity was operationalised by having the accounting ratios computed from financial statements closer to corporate failure (one year before corporate failure) while high task complexity involved computing ratios one year further away from corporate failure (i.e., two years before corporate failure). This operationalisation of task complexity was adopted because previous studies found that it was easier to predict corporate failure (the criterion event) using accounting ratios (cues) closer to the criterion.

Simnett (1996) manipulated task complexity by varying the information load (a component complexity dimension of Wood's (1986) general task complexity model) in a bankruptcy prediction task. Specifically, subjects in the low information load group received twelve pieces of information while high information load subjects had 30 items. It was found, that both information load and subjects' experience affected judgment accuracy, consensus and processing strategies, especially in the case where subjects were allowed to choose their information set.

2.2.2.2 Audit Studies Addressing Task Complexity and Feedback

A range of psychology studies have investigated the impact of task complexity on different types of feedback. For example, Brehmer (1969), Dean et al. (1972), Schmitt et al. (1976), and Steinmann (1976), as discussed in Section 2.1.

In auditing, the only study that has examined the impact of task complexity on feedback is Hirst, Luckett and Trotman (1999). In this study, task complexity was manipulated by varying task predictability at two levels in a bankruptcy prediction task involving OFB, TPF, and OFB+TPF. It was found that generally OFB, TPF, and OFB+TPF promoted judgment accuracy. Also, in both task complexity levels, OFB, and OFB+TPF promoted judgment accuracy over time, but no such effect was found in respect of TPF. In the low complexity condition (high in predictability), the judgment performance of the OFB group was significantly higher than that of the TPF group.

CHAPTER 3

HYPOTHESIS DEVELOPMENT

As outlined in the literature review, there are studies investigating the effects of various types of feedback on improving judgment performance in both psychology and accounting literatures. Although most psychology studies have generic tasks and most accounting studies have non-auditing tasks (such as bond rating and performance evaluation), their results provide a guideline for developing testable hypotheses in this study. This chapter consists of three main sections, namely, feedback, self-insight, and the transfer of learning. In each section, hypotheses are developed relating to the respective issues.

3.1 Feedback

Five feedback conditions were manipulated across two levels of task complexity in this study, and their respective effect in improving judgment performance was evaluated. The feedback conditions are NFB, OFB, TPF, CFB, and CombFB. One level of task complexity required configural information processing and thus was high in task complexity and the other was a low complexity task, which required linear additive information processing.

3.1.1 Outcome Feedback (OFB)

It has been found in both psychology and accounting literatures that OFB can be useful in improving judgment performance (Adelman 1981; Schmitt et al. 1977) when the task satisfies the following three requirements (see, for example, Hirst and Luckett 1992 for review):

- it is highly predictable;

- it is relatively simple; and
- the relationship between the cues and the criterion is largely linear.

In addition to the above task characteristics, individuals likely to benefit from OFB are those who have high initial domain- and task-specific knowledge and are familiar with the task. Individuals would find it easier to develop a stable and appropriate judgment strategy from the OFB for a relatively simple and familiar task where the cues are linearly related to the outcome or criterion. Thus, based on OFB, individuals can fine-tune their judgment models and cue weights to improve their judgment performance in a task satisfying the above-mentioned characteristics.

On the other hand, there are studies in the psychology literature which found that OFB is ineffective in improving judgment performance (Balzer et al. 1989; Einhorn and Hogarth 1981), especially in complex inference tasks even after extensive training (Goldberg 1968). In an accounting study relating to bond rating, Kessler and R. Ashton (1981) find no effect on judgment performance after their student subjects received OFB. Judgment performance may even be adversely affected by OFB in an experiment with 40 student subjects and a task involving a 3-cue generic task (Holzworth and Doherty 1976).

One of the factors that increases the complexity of a task is the functional form relating the criterion to the cues. Brehmer (1969) finds that the learning of subjects (40 undergraduate students) is slower in a task involving two configural cues than a task of two linear additive cues in the first few blocks with a total of 200 trials. Deane et al. (1972, Experiment 1) provide similar results in an experiment with 60 undergraduates and find that judgment performance is higher for linear tasks than for non-linear tasks after receiving OFB. This suggests the following hypothesis:

H1a OFB is more effective in improving judgment performance in a non-configural task than a configural task.

3.1.2 Task properties Feedback (TPF)

Most studies in both psychology and accounting literatures found TPF effective in improving judgment (see, for example, Balzer et al. 1989; Tuttle and Stocks 1998) in a wide variety of tasks, ranging from some simple generic tasks to complex tasks, such as bond rating (Kessler and R. Ashton 1981), performance evaluation in professional settings (Luckett and Hirst 1989), and forecasting (Remus et al. 1996). The results are robust across tasks in terms of the formal characteristics of tasks (such as the number of cues and their relationship to the criterion) and the substantive characteristics of tasks (such as the content of the task). Balzer et al. (1989) after extensive review of the feedback literature conclude that TPF is more effective in tasks demanding more cognitive activities and state that "the evidence suggests that it is the TI component ... that has made it work" (p.428). The "TI component" is equivalent to TPF information in the present study. Doherty and Balzer (1988) reach a similar conclusion in their review paper.

TPF from external sources benefits performance by providing instruction that guides the performer to correct responses. Performance gains from TPF are likely due to its guidance function (Salmoni et al. 1984; Schmidt 1991). From the TPF, subjects would know which cues are valid (and to what extent each is valid) and the functional form between the cues and the criterion. They can quickly put that information to use in making their judgments. This suggests the following hypothesis:

H1b: TPF is effective in improving judgment performance in tasks involving both configural and non-configural processing.

3.1.3. Cognitive Feedback (CFB)

In the psychology literature, it is generally found that individuals' descriptions of their judgment policies are often inaccurate and that they have difficulties externalising their policies. This is particularly the case if the information processing involves correlated cues (Reilly and Doherty 1992) and configural cues. Schmitt et al. (1976) found no main effect for CFB on judgment achievement for tasks of relatively high predictability. In an experiment designed to investigate the relative effectiveness of CFB, Balzer et al. (1994) found that CFB did not improve judgment performance in a highly predictable baseball judgment task involving five linear cues of low intercorrelation. Thus, the provision of CFB in a simple task, which normally involves a linear combination of a small number of cues, is likely to lead to little improvement in judgment performance because the subjects may have a reasonable knowledge of their own judgment policies.

After an extensive review of the psychology literature on feedback, Balzer et al. (1989) suggested that research on the effectiveness of CFB should not be abandoned even though most studies indicated that it was not as effective as TPF in improving judgment performance. They suggest that in a complex task individuals may not be aware of which cues are important to their judgments and are unaware of the functional forms relating their judgments to each cue (Balzer et al., 1989). Therefore, the provision of CFB would increase their awareness of their own judgment policies. As a result, CFB would be more effective in improving their judgment performance. They noted that

> "Many laboratory tasks investigating CI have been at the low end of the dimensions of complexity; hence, whether CI may facilitate performance in more complex tasks is not settled. If one assumes that many interesting tasks in the world include a large number of cues, non-linear cue-criterion relationships,

and complex aggregation rules, then the study of CI in more complex environments seems prudent" (p. 428).

They used the term CI to represent cognitive feedback information. Balzer et al. (1992) speculated that CFB may be effective in tasks of high complexity using curvilinear functional forms, in integrating cues, or in a configural environment, although they find no significant difference in judgment performance between the CFB and NFB groups in their experiment. They noted that while CFB "was not necessary to improve judgment performance in the present task, perhaps it would be beneficial in more 'complicated' tasks with intercorrelated cues, non-linear relationships between cues and criterion, environments that are highly configural, and so on" (p. 51). This suggests the following hypothesis:

H1c CFB is more effective in improving judgment performance in a task that does involve configural processing than in one that does not.

3.1.4 Combined Feedback (CombFB)

Previous studies have found no advantage for providing CombFB over TPF alone (Balzer et al. 1992; Balzer et al, 1994; Nystedt and Magnusson, 1973; Steinmann, 1974). One explanation is that if subjects are aware of their judgment policies then TPF alone would allow them to compensate for inappropriate weighting strategies, particularly in low task complexity cases in which cue weights differ substantially from one another. Despite the above findings on the ineffectiveness of CombFB over TPF alone, Balzer et al. (1992) suggested that CFB may be more effective in more complex tasks and environments that are highly configural. Thus, TPF together with CFB could improve judgment performance in a task of high complexity and configurality. Hoffman et al. (1981) also suggested that TPF may be sufficient for high-level performance for some tasks, but CombFB may be required in tasks of greater complexity (p.80). This suggests the following hypotheses:

H1d CombFB leads to greater improvement in judgment performance for a configural task than a non-configural task.

H1e The improvement in judgment performance between CombFB and TPF is greater for a configural processing task than for a non-configural processing task.

3.2 Self-insight

A potentially important variable, namely, self-insight has not been examined in feedback research. This is surprising because one of the reasons for CFB being developed was due to evidence suggesting that individuals lacked self-insight, and they exhibited a poor understanding or implementation of judgment policy (Balzer et al. 1989). Subjects often describe their policies inaccurately (Balke et al. 1973; Reilly and Doherty 1992; D. Summers et al. 1970), or they overestimate the number of cues that are important to them (McKenzie 1997; Slovic and Lichtenstein 1971). Also, they overestimate the weights of those cues that are not important to their judgments (Slovic 1969). The general research findings suggest that individuals' lack of insight into their judgment policies provides a strong rationale for developing CFB for individuals despite the costs and time involved in the process (Balzer et al. 1989). The self-insight of subjects in accounting studies tends to be higher. R. Ashton (1974) had auditors as subjects and found high self-insight (mean subjective/objective cue weight correlation of 0.89) in an internal control assessment task. Joyce (1976) also reported high self-insight with auditors as subjects with a mean subjective/objective cue weight correlation of 0.53. W. Wright (1977) found MBA students had a fair level of self-insight (correlation of subjective/objective cue weight of 0.388) in a stock price prediction task involving seven cues. Also, student subjects in a financial distress prediction task exhibited a fair amount of self-insight (Tuttle and Stocks 1997). However, Luckett and Hirst (1989) found auditor subjects who exhibited poor self-insight in a performance evaluation task.

3.2.1 Self-insight and CFB

Self-insight of an individual can moderate the effectiveness of CFB in improving judgment. Recall that CFB provides feedback to subjects about their functional forms for combining cues and cue weights. Subjects with perfect self-insight know their own functional forms and cue weights in making judgments even without the CFB. Therefore, CFB would be of virtually no value to subjects with perfect self-insight. However, very few subjects would have perfect self-insight even in a reasonably simple task (Balzer et al. 1989; Luckett and Hirst 1989; McKenzie 1997; Reilly and Doherty 1992; Tuttle and Stocks 1997).

On the other hand, subjects with poor or low self-insight may not know the functional form or weights they put on each cue in their models. For this group of subjects, CFB provides them with additional information, which is related to their judgment models and cue weights. By comparing their own domain- and task-specific knowledge and their mental judgment models and cue weights, they are in a better position to make appropriate adjustments based on the information in the CFB. This suggests the following hypothesis:

H2a The effect of CFB in improving judgment performance would be higher for subjects with low self-insight than for those with high self-insight in tasks of both low and high complexity.

3.2.2 Self-insight and CombFB

Previous studies also investigated the effect of CombFB (TPF and CFB) on judgment performance. The results tended to suggest that CombFB did not improve the judgment performance of subjects over and above that achieved by TPF alone (see, for example, Balzer et al. 1989 for review; Galbraith 1984). These results are not surprising if subjects given CombFB were of high self-insight, as the CFB component of the CombFB would provide them with little additional information as they would have knowledge of their own judgment models and cue weights.

On the other hand, CombFB is likely to be more effective in improving judgment performance than TPF for subjects with low self-insight. For this group of subjects, they would gain more knowledge of their judgment models and cue weights from the CFB component of the CombFB. They have the opportunity to improve their judgment performance by comparing the most appropriate judgment model and cue weights (from the TPF component) with their own (from the CFB component). By knowing the difference between their models and cue weights from the TPF (the benchmark), the CFB component facilitates the adjustment of their own models and cue weights to be in line with those of the TPF. Based on the above discussion, the following hypothesis can be derived:

H2b The effect of CombFB in improving judgment performance would be higher than the effect of TPF alone for subjects with low self-insight in tasks both of low and high complexity[1].

3.3 Transfer of Learning

One aspect of learning in professional settings is the transfer of learning (or inter-task learning). That is, the learning from one task enhances performance in a somewhat different context / task (Salomon and Perkins 1989). While Chapter 2 outlined the importance of being able to transfer acquired knowledge and cognitive skills to new tasks and problem situations, a comprehensive review of the feedback literature by Balzer et al. (1989) found only one previous study (Steinmann 1974) that investigated whether OFB, TPF and CombFB on a particular judgment task could be transferred to other tasks. McCarthty et al. (1995) made a similar comment after reviewing the educational psychology literature that "Minimal research related to feedback and transfer has been published…" (p. 142).

As noted in Chapter 2 it is generally accepted that "far" transfer (such as transferring / applying knowledge to a less similar task or concept) is more difficult to obtain than "near" transfer (De Corte 1996). This notion of transfer distance relates to the degree to which the transfer task is remote or novel in comparison with the original learning / training task. I would expect subjects practiced in a task of low complexity and receiving various kinds of feedback would not acquire the knowledge required to perform a task of high complexity, such as one involving configural cue processing.

[1] Subjects were classified as having high or low self-insight according to the median of all subjects who completed Task 1A or Task 1B.

In the auditing context, experience and knowledge gained in a task involving configural cue processing would be easier to transfer to another audit task also involving an interaction of cues. Subjects completing a task of high complexity (Task 1A) would learn better and be in a better position to transfer the knowledge to another configural task (Task 2). Based on the above, the following hypothesis is derived:

H3a Subjects having practised in a configural cue processing task (Task 1A) would perform better in a subsequent task (Task 2) requiring similar information processing than those who practised in a non-configural task (Task 1B).

Studies in psychology indicate that learning and the transfer of learning to novel contexts requires the explicit rules of the task to be made known to the individuals (Renkl 1999). TPF provides this information as it states the functional form and cue weights to complete a task. Subjects receiving TPF and having practised a complex and similar task would learn the configural nature of cues. When they are required to perform tasks with a similar configural cue pattern, or a different pattern, they would find it easier to transfer the knowledge and apply it to the new task. If individuals are provided with CFB and complete a complex task (Task 1A), they can improve their performance by knowing their own judgment policies. However, the learning may or may not be effective. There is no guarantee that their judgment policies are the most appropriate ones. Steinmann (1974) found that subjects trained with CFB were able to transfer learning to other tasks.

Subjects receiving CombFB could exhibit learning and be able to transfer the knowledge to another task. They not only know the most appropriate judgment policy, but also their own policies. They should be aware of the discrepancies and make adjustments. Further, they should be more ready to adjust and internalise the correct judgment policy and to put these pieces of information into practice. With the internalisation and practice of a configural cue processing strategy they can transfer it and apply it to subsequent audit tasks, which require a similar or different form of configural cue processing.

Subjects in the NFB groups receive no feedback. The learning would be minimal, especially in a task of high complexity. Thus, their performance in a subsequent task should be lower than that in TPF, CFB, or CombFB groups.

H3b Subjects who completed a task of high complexity and received TPF perform better in Task 2 than subjects in the NFB groups.

H3c Subjects who completed a task of high complexity and received CFB perform better in Task 2 than subjects in the NFB groups.

H3d Subjects who completed a task of high complexity and received CombFB perform better in Task 2 than subjects in the NFB groups.

CHAPTER 4

RESEARCH METHODS

A controlled laboratory experiment was developed, pilot tested and conducted in order to collect the data to test the hypotheses developed in Chapter 3. A laboratory experiment is the appropriate approach as it allows effective manipulation of the variables being studied. Also, controlled laboratory experimental settings allow researchers the ability to abstract and control other potentially influential variables (Libby and Luft, 1993). The details of the experimental design, task, subjects, and procedures are described in the following sections.

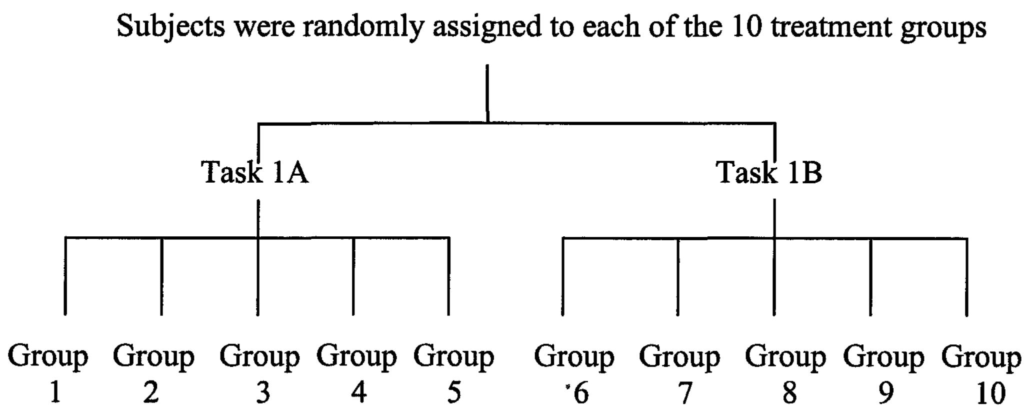
4.1 Experimental Design

The experiment was a 5 (feedback treatments) x 2 (task complexity) x 2 (blocks) full factorial design. Feedback and task complexity were manipulated between-subjects and block was a within-subjects variable. After completing the above experiment, subjects were required to complete another risk assessment task (Task 2, details of which will be outlined later). This task was to examine the transfer of learning. A diagrammatic presentation of the main research design is depicted in Exhibit 4.1. Five feedback treatments were manipulated in the experiment, namely, no feedback (NFB), outcome feedback (OFB), task properties feedback (TPF), cognitive feedback (CFB), and combined feedback (CombFB). There were two groups of subjects (groups 1 and 6) who received no feedback. They served as control groups. OFB was provided to the outcome feedback groups (groups 2 and

7) after each case in the first block of 16 trials. For subjects in the TPF, CFB, and CombFB groups, the relevant feedback was presented to them after the first block, but before the second blocks of trials. TPF was provided to the TPF (groups 3 and 8), and CFB was provided to CFB groups (groups 4 and 9). Both TPF and CFB were provided to the CombFB groups (groups 5 and 10).

Exhibit 4.1

Diagrammatic Presentation of the Research Design



Notes:

1. Groups 1 and 6 received no feedback (NFB);
   Groups 2 and 7 received outcome feedback (OFB);
   Groups 3 and 8 received task properties feedback (TPF);
   Groups 4 and 9 received cognitive feedback (CFB); and
   Groups 5 and 10 received combined feedback (CombFB).

2. All subjects in each group completed the following activities:

i. Practised a dummy task on a payroll system to let them familarise themselves with the operations of the computer program;
ii. Completed the first block of 16 cases of the respective task (only subjects in the OFB groups received OFB after they completed each case);
iii. Answered questions on demographic data and self-insight, and completed a filler task on audit independence;
iv. Subjects in the TPF, CFB and CombFB groups were given relevant feedback at this stage;
v. Completed the second block of 16 cases of the respective task; and
vi. Completed Task 2.

Task complexity was manipulated across two tasks (Task 1A and Task 1B). Task 1A was of high task complexity and Task 1B was of low task complexity. For each task, there were 2 blocks, each of 16 cases, representing two one-half fractional replications of a 2^5 factorial design. Each cue was manipulated at 2 levels. Specifically, the two one-half fractional replications of 16 cases were administered respectively in Block 1 and Block 2 of the experiment.

4.2 Tasks

Two audit tasks were employed in this experiment. They differed in terms of the level of complexity required to process the information in the respective task. The two tasks are described in detail in the following sections.

4.2.1 Task 1A

Task 1A was adapted from Brown and Solomon (1991, Experiment 1). It consisted of a hypothetical audit case involving judgment of the risk of misstatement of an account balance based on part of an audit program for accounts receivable. The case included background information and specific information concerning audit procedures undertaken (see Exhibit 4.2). It should be noted that the original task in Brown and Solomon (1991) has six cues, one of which was kept constant. This constant cue was not included in the present study. Instead, the information content of this constant cue was incorporated in the background information of the audit case. This was necessary because subjects were required to rate the importance of each of the cues after they had completed the first block of 16 trials. Their ratings were then compared with the statistically modeled weights based on their judgments in the first block, to assess their self-insight. Including a constant cue in the task, which would not account for any variance in the subjects' statistical models, would result in subjects assigning a certain weight to that cue when they were asked to externalise the importance of each cue in their judgments. This would unfairly understate their self- insight. Therefore, it was decided to include only the five manipulated cues in the Brown and Solomon's (1991, Experiment 1) task.

The subjects were told that they would be given a series of checklists (completed by one of their staff auditors), each representing a partially completed portion of the accounts receivable audit program. They were then asked to assess the misstatement risk of the account balance based on the audit evidence-to-date. A sample of the checklist and judgment scale is depicted in Exhibit 4.3. Appendix A shows the computer screens with the instructions and background information of Task 1A.

Exhibit 4.2

Background Information and Specific Information on Auditing Procedures of the Audit Case (Task 1A)

Assume you are a senior-level auditor and that one of your clients is Tritronic Ltd. Tritronic, a large designer and manufacturer of home electronic appliances, is a privately held company. The company has its design and trading activities in Hong Kong and production is in Shenzhen, China. The company has not presented significant auditing problems during your firm's five-year tenure as its external auditor, and has maintained moderately good internal accounting control systems. During the past two years, Tritronic has been computerizing and integrating its accounting, management and manufacturing information systems.

Currently, you are performing Tritronic's 1999 year-end audit (for the accounting year from 1 July 1998 to 30 June 1999) and are evaluating evidence provided by the audit of accounts receivable. During engagement planning, the inherent risk that a material error could arise in accounts receivable was assessed to be moderate for both existence and valuation. Analytical procedures had been conducted to verify the adequacy of the allowance for bad debts to compare the current year's number of days accounts outstanding with that of previous years, and to compare the allowance as a percentage of accounts receivable with that of previous years. The analytical procedures were completed with no exceptions noted.

For (16) randomly ordered cases, you will be presented with the results from a portion of the accounts receivable audit program performed by your staff auditors. For each case, you will be asked to assess the risk that accounts receivable, given the audit evidence-to-date, could be materially misstated AS A RESULT OF recorded accounts NOT EXISTING at year-end.

Additional accounts receivable audit procedures information:

A. The sample for accounts receivable confirmation was drawn to obtain 95 percent reliability estimates. WHEN indicated as being COMPLETED, responses have been received (through second and third requests) from 100 percent of the sample.

B. Verification of accounts receivable sales transactions and observation of subsequent collections is a SEPARATE audit procedure from the confirmation of those accounts. The sample for this procedure (also drawn to obtain 95 percent reliability estimates) is INDEPENDENT of that for the confirmations.

C. The aged accounts receivable listing is reviewed for material receivables from affiliates, directors, and other related parties; debentures and long-term receivables; and accounts with significant credit balances.

Exhibit 4.3

A Sample of the Checklist and Response Scale of the Accounts Receivable Case (Task 1A)

	Procedure	Results
A.	Confirm a 95% reliability sample of year-end A/R (positive form).	Not Completed as of This Date
B.	Verify sales transactions for and observe collections of a 95% reliability sample of year-end A/R.	Completed With No Exceptions Noted
C.	Reconcile A/R subsidiary ledger to the general ledger control account.	Completed With No Exceptions Noted
D.	Review aged A/R listing for related party transactions, long-term debentures, etc.	Not Completed as of This Date
E.	Review Board of Directors' meeting minutes and correspondence files to determine if receivables were factored or discounted.	Not Completed as of This Date

Given the audit evidence as represented by the partially completed audit program segment (above), assess the risk that the year-end accounts receivable balance could be materially misstated as a result of recorded accounts NOT EXISTING.

0 10 20 30 40 50 60 70 80 90 100

NO RISK — MAXIMUM RISK

Confirmation of accounts receivable (Procedure A) and verification of sales transactions and observation of subsequent collections (Procedure B) were intended to be substitutable audit procedures with respect to the existence assertion. Procedures C through to E were intended to be accounts receivable audit procedures less directly related to the existence assertion. All five audit procedures (A, B, C, D, and E) were factorially manipulated at two levels each (either "completed with no exceptions noted" or "not completed as of this date").

In this task, auditor subjects were required to make a risk assessment relating to the *existence* assertion of the accounts receivable balance. An effective and commonly used procedure would be confirmation of individual account balances with debtors. If the confirmation requests were returned from debtors without exceptions noted, the auditor might assess that the existence assertion of the accounts receivable balance is valid. Thus, the risk of misstatement would be low. However, if a sufficient number of confirmation requests were not returned or returned with exceptions noted, the auditor may have to expand his/her substantive audit procedures, such as via verification of sales transactions and observation of subsequent collection, to collect further evidence. In other words, confirmation can be considered as the primary audit procedure and verification of sales transactions and observation of subsequent collection may be viewed as alternative or substitute procedures.

This substitution effect, as specified by the domain- and task-specific knowledge and audit practices, requires auditors to evaluate the results of audit procedures in a particular pattern or form of configuration – a compensatory-form ordinal interaction (Brown and Solomon 1991). That is, auditors need to recognise that the two tests are substitutable, with the existence of one compensating for the absence of the other. The requirement to configurally process evidence (cues) makes the judgment more complex than a task which only requires linear processing of evidence (cues), such as Task 1B (see discussion below).

4.2.2 Task 1B

Task 1B is similar to Task 1A, with only one difference – Procedure B "verify sales transactions for and observe collections of a 95% reliability sample of year end A/R" is replaced by another audit procedure "verify adequacy of allowance for bad debts using analytical procedures". This new audit procedure is a less direct test of the existence assertion and is not intended to be substitutable for Procedure A. Based on domain- and task-specific knowledge and audit practices, Task 1B only requires auditors to make a risk assessment by a linear combination of audit evidence (cues). This task is easier to complete and thus has a lower level of task complexity, when compared with Task 1A.

Because of the new cue on the conduct of analytical procedures in Task 1B, the information relating to the conduct of analytical procedures as stated in the background information of Task 1A differs in one aspect. Specifically, the last two sentences in paragraph two of the background information of Task 1A (relating to the conduct of the analytical procedures) were deleted (see Exhibit 4.2).

4.2.3 Task 2

In order to assess whether the learning in Task 1A and Task 1B under different feedback conditions can be transferred to another task, subjects were required to complete Task 2. Task 2 was adapted from Brown and Solomon (1990). It involves two forms of configural cue processing, namely, compensatory-form and amplifying interactions (discussed below). It is a hypothetical case on the control risk assessment of the cash disbursements procedures of an audit client. The case included background information and described the client's cash disbursements controls. Subjects were presented with a series of checklists of cash disbursement controls completed by a staff auditor and asked to assess the risk that cash disbursements could be materially misstated as a result of cheques being written and / or disbursed for improper purposes. Exhibit 4.4 shows an example of the cash disbursements control checklists.

Exhibit 4.4

A Sample of the Instructions and Details of the Internal Controls over Cash Disbursements Case (Task 2)

Assume you are a senior-level auditor and that one of your clients is Nortack, Inc. Nortack, a large processor and merchandiser of agricultural commodities, is a privately held company that has debt covenants requiring audited financial statements prepared in according with GAAP. The company has not presented significant auditing problems during your firm's five-year tenure as its public auditor. Nortack's management is actively involved both in designing the company's internal controls, as well as reviewing existing internal controls. The employees who administer Nortack's internal controls are well trained and supervised, with clearly defined responsibilities. Nortack has a relatively autonomous internal audit department that is adequately staffed and supervised; the department head was a manager for a Big-5 CPA firm, and most of the internal auditors have CPA certificates. During the past five years, Nortack has been computerising its accounting and information systems.

Currently, you are planning Nortack's 1998 audit engagement and are evaluating its internal controls to determine the extent to which you will rely on them in planning the year-end audit work. For 32 randomly ordered cases, you will be presented with a portion of a cash disbursement internal control questionnaire completed by an auditor on your staff. For each case, you will be asked to assess the risk that the specific controls could give rise to a material misstatement of cash disbursements AS A RESULT OF checks being written and / or disbursed for improper (unauthorised and / or invalid) purposes.

Additional cash disbursement internal controls information:

A. The authorization for approving expenditure requests has been designated by the Board of Directors at various management levels, depending upon the nature and amount of the request. Expenditure authorisation is indicated on purchase orders.

B. The cash disbursement department has the responsibility for verifying the propriety of expenditures and for recording them in the voucher register. The original copy of the voucher has a copy of the vendor's invoice, receiving report and purchase order attached.

C. Primary check signers carefully scrutinize vouchers and supporting documentation at the time checks are signed.

D. WHEN THEY EXIST, second check signers are independent of all other expenditure and cash disbursement functions

Exhibit 4.4 (Cont'd)

	CONTROL PROCEDURE	YES	NO
A.	Are protective writing devices used to inscribe amounts on checks?	✓	
B.	Are properly approved vouchers required for check preparation?	✓	
C.	Are all check signers designated by the Board of Directors?	✓	
D	Are the primary check signers independent of:		
	1. Purchasing and those requesting expenditures?	✓	
	2. Persons approving vouchers?		✓
	3. Persons processing and recording cash disbursements?		✓
E.	Is an independent second check signer required who carefully scrutinises the supporting documentation?	✓	
F.	Does internal audit investigate payments made to payees not on an independently approved payee listing?	✓	

Given the controls as requested above, assess the RISK that cash disbursements could be materially misstated AS A RESULT OF checks being written and / or disbursed for improper (unauthorised and / or invalid) purposes.

0 10 20 30 40 50 60 70 80 90 100

NO RISK ... MAXIMUM RISK

Procedure D in the cash disbursements controls checklist has three related separation-of-cash-disbursements-duties controls. These are the primary separation-of-duties controls in the cash disbursements function. The assessed risk that cash disbursements could be materially misstated as a result of cheques being written and / or disbursed for improper (unauthorised and / or invalid) purposes would be low if Procedure D was present and operating effectively. Procedure E is a secondary separation-of-cash-disbursements control. It is a preventive control with respect to the control objective that cash disbursements are for proper purposes. Procedures D and E are supposed to compensate for each other. That is, the risk that cash disbursements are for improper purposes would be low if either one of the two procedures is present or both are present. However, if the two procedures were absent, the risk would be assessed as substantially higher. Procedure F is primarily a detective cash disbursements control, although it may have indirect preventive effects. Procedure F would amplify the effect of Procedure D. If Procedure D is present and operating effectively, control risk would be assessed as relatively low, but if Procedure F was also present and operating effectively, control risk would be assessed to be even lower. Alternatively, if Procedure D was not present, control risk would be assessed to be high regardless of whether Procedure F was present or not present (Brown and Solomon 1990).

Procedures A through C are cash disbursements controls not related primarily to the stated cash disbursements control objectives (or, at least, not related as strongly as Procedures D, E, and F). These control procedures, therefore, are not expected to compensate or amplify Procedure D to the same extent as those in Procedures E and F.

Five of the six control procedures (A, C, D2 and D3 jointly, E, and F in Exhibit 4.4) were manipulated at two levels each (Yes or No), and two control procedures (B and D1 in Exhibit 4.4) were held constant (Yes) in all cases.

4.3 Subjects

Subjects for this study were 236 Big-5 audit seniors with two to five years of audit experience in Hong Kong. Most of them graduated from the university where I am teaching. They were personally contacted and invited to participate in the experiment. They volunteered to participate in the study for a souvenir (a plastic watch with the logo of the Department of Accountancy) and ten lucky draw cash prizes (about A$250 or US$130 each). Once they agreed to participate and booked a time to come to conduct the experiment, they were randomly assigned to one of the 10 treatment groups.

Of the 236 auditor subjects, the responses from five subjects were not usable because of diskette problems and only part of the results were recorded or the data could not be retrieved. In addition, the responses of seven subjects were excluded because they failed the manipulation check in recalling and ranking the three most important cues in the feedback presented to them. As a result, there were 224 valid responses (113 female and 111 male auditors). Table 4.1 presents the distribution of firm affiliation of auditors.

Table 4.1
Firm Affiliation of Auditor Subjects

Firm Affiliation	Number of Auditors	Percentage
Arthur Andersen	44	19.6
Deloitte Touche Tohmatsu	66	29.5
Ernst and Young	33	14.7
KPMG	39	17.4
PricewaterhouseCoopers	42	18.8
Total	**224**	**100%**

The distribution of auditor subjects from each of the Big-5 firms is not equal. It is dependent on their respective size of operations and on the schedule of work of each audit firm when the experiment was conducted. At that time, KPMG, for example, had a few special assignments and many of their auditors were out-posted to China.

Table 4.2 shows the distribution of audit experience of subjects. No firm effect or gender effects were found to be significant.

Table 4.2
Audit Experience of Subjects

Year	Number of Auditors	Percentage
2	71	31.7
3	114	50.9
4	29	12.9
5 or above	10	4.5
Total	**224**	**100%**

The average audit experience of subjects is 2.9 years. Most of them (about 70%) had at least 3 years audit experience, and they would be familiar with the risk assessments in the experiment. Post-experiment discussion with them also confirmed this. Table 4.3 presents their self-reported proportion of audit assignments on which they are required to make A/R and cash disbursements internal control risk assessments.

Table 4.3
Self-reported Experience of Risk Assessment in Audit Assignments

Proportion	Number of Auditors	Percentage
Above 70%	90	40.2
51 – 70%	83	37.1
31 – 50%	38	17.0
Less than or equal to 30%	13	5.8

Most of the auditor subjects (91%) possess professional accounting qualifications and the majority of them (91%) had a major in Accounting while another 5% were business graduates.

4.4 Procedures

Subjects were randomly assigned to one of the ten treatment groups when they agreed to participate in the experiment and each subject was asked to select a time, which was convenient for him / her to come to the laboratory.

Subjects came at the scheduled time and completed the experiment in one of the two computer rooms (vacant staff offices converted temporarily to computer rooms for this experiment). There were two personal computers in each of the rooms. The study was programmed using IBM personal computers to standardise instructions. Only one to two subjects were scheduled for each session to allow for machine malfunction and to leave ample space for subjects to take notes and to prevent them from seeing another participant's computer screen and notes. They sat at the personal computers and worked individually at their own pace, clicking the mouse or typing responses to cases that appeared on the computer screen. They were told that there was no time limit, but that the computer would record the time they spent on the experiment. They were permitted to ask questions during the experiment. There was a piece of paper outlining the stages of the experiment on the desk so that it would be easier for subjects to follow.

All subjects (with the exception of those in the no feedback groups) completed six stages: (1) trial stage; (2) first block of 16 cases of Task 1A or Task 1B; (3) self-insight and demographic data questions and a filler task; (4) second block of 16 cases of Task 1A or Task 1B; (5) manipulation checks and questions on their perception of the usefulness of the feedback provided in feedback groups; and (6) a block of 16 cases of Task 2. Subjects in the two no feedback groups did not have the fifth stage as they did not receive any feedback information. The details of each stage are outlined below.

Stage 1 aimed at familiarising subjects with the operations of the computer system. It began with brief instructions on the personal computer and was followed by an audit procedure evaluation case involving payroll. Each subject evaluated the risk involved in 2 payroll cases. These 2 trials were not related to subsequent experimental tasks in the study. However, the presentation and action required from subjects was similar to those in subsequent tasks. There were built-in mechanisms to remind subjects if they clicked on the wrong side of the response scale. One of the two cases had a low level of risk because most of the audit procedures were completed with no exception noted. If subjects clicked on the side of the response scale indicating a high level of risk (50 or above in a scale of 0 to 100), the computer would display a warning message ("Did you click on the wrong side of the scale? Most auditors would assess the risk at a low level in this case") indicating that he/she might have misunderstood the scale. Another case had a high level of risk because most of the audit procedures were not completed. If subjects' risk assessment was on the low side of the scale (40 or below on a scale of 0 to 100), the computer would also display a similar warning message on the screen.

In stage 2 subjects completed the first block of 16 cases of Task 1A or Task 1B depending on the treatment group. Subjects in groups 1 to 5 had Task 1A and those in groups 6 to 10 had Task 1B. During the experiment, they read the instructions on the screen and clicked on their selected response (i.e., risk assessment) to each case. Subjects were permitted to ask questions at any time during the experiment. Because subjects went through the practice phase and were familiar with the computer system, and the instructions were clear, few questions were asked. Only subjects in groups 2 and 7 received OFB after each of the 16 cases in the first block. Subjects in the other treatment groups did not receive any type of feedback at that stage.

Stage 3 solicited subjects' self-insight into their own judgment models in the first block of trials. A message on the screen then asked them to take the diskette from the computer and give it in to the researcher. They were then given another diskette and a 5-page filler task (explained below). They then started to answer questions soliciting their self-insight and demographic data. That is, they had to externalise the subjective weight they put on each cue in making the risk judgments by rating the importance of each cue from 0 (not important at all) to 100 (very important). An example of the cue-importance rating instrument is shown in Exhibit 4.5. Subjects' demographic data were also solicited in this stage. While they were completing this part, I modeled their judgment policies with their responses in the first block of 16 cases, to provide cognitive feedback to subjects belonging to the cognitive feedback and combined feedback groups. In order to standardise the treatment for all groups, all subjects were required to take the diskette from the computer, hand it to the researcher, and swap it for another diskette. Also, before they commenced Stage 4, they were asked to complete a filler task so that I had sufficient time (normally about 15 minutes) to model subjects' judgment policies for the provision of CFB and CombFB. The filler task was a case study asking subjects to assess the extent of auditor independence. It was not related to this research. The major purpose of having this filler task was to keep subjects occupied when I was preparing the cognitive feedback for some of the treatment groups.

Exhibit 4.5

Instrument to Solicit Self-insight

The importance of each audit procedure:

Please rate the importance of each of the following audit procedures when you are making the A/R risk assessments for the above 16 cases by TYPING A NUMBER FROM 0 TO 100. You may assign any number from 0 to 100 to indicate the importance. A rating of 100 indicates that the audit procedure is very important in your A/R risk assessments. On the other hand, a rating of 0 indicates that the audit procedure is not important at all in your A/R risk assessments.

Confirm a 95% reliability sample of year-end A/R (positive form). (0 – 100) ______

Verify sales transactions for and observe collections of a 95% reliability sample of year-end A/R. (0 – 100) ______

Reconcile A/R subsidiary ledger to the general ledger control account. (0 – 100) ______

Review aged A/R listing for related party transactions, long-term debentures, etc. (0 – 100) ______

Review Board of Directors' meeting minutes and correspondence files to determine if receivables were factored or discounted. (0 – 100) ______

Stage 4 had the second block of 16 cases of Task 1A or Task 1B. TPF was provided to groups 3 and 8, CFB was provided to groups 4 and 9, and CombFB was provided to groups 5 and 10, *before* they started this block of trials. After subjects in these six groups received and studied the feedback, they continued with the rest of the experiment. There was no further outcome feedback provided to the outcome feedback groups (groups 2 and 7) in this stage.

Stage 5 solicited subjects' opinions on the usefulness, current helpfulness, and future helpfulness of the feedback provided to them for treatment groups receiving various types of feedback. Subjects in the no feedback groups did not complete this stage. The scales were validated and used in Balzer et al. (1992) and Balzer et al. (1994). There were 2, 8, and 2 statements on the usefulness, current helpfulness, and future helpfulness of the feedback, respectively. Subjects responded to a 5-point Likert scale for all 12 statements.

Also, subjects receiving TPF and CFB were asked to recall the feedback provided to them at the end of Stage 3, and to rank the importance attached to the three most important audit procedures by the partner (themselves). The ranking was simple. (1 = most important; 2 = second important; 3 = third important). Subjects receiving CombFB were asked to recall and rank both the partner's and their own importance attached to the three most important audit procedures. The information collected served as a manipulation check to make sure that the subjects understood and remembered the feedback presented to them. Subjects in the NFB and OFB groups were not required to complete this part as they did not receive any of the information.

Stage 6 incorporated the 16 cases of Task 2. The cases were the same for all subjects. When they completed the trials, they were thanked, debriefed, and given the souvenir and the details of the lucky draw.

4.5 Presentation of Feedback Information

To provide feedback and to evaluate performance, the following information was obtained. For Task 1A, I asked five Big-5 partners to complete the task (32 cases) and one of the partners' judgments were selected as the benchmark. This partner's judgments were selected because (1) his judgments were close to the average of all the partners, and (2) he exhibited the most configural cue processing in the desired direction among the audit partners. For Task 1B, I asked another three Big-5 partners to complete the task (32 cases) and I chose the partner's judgments that were closest to the average of three of them. Again, this partner's judgments were used as the benchmark to compare with the judgments of subjects completing this task.

OFB was provided to subjects in the OFB treatment group after each case in the first block of 16 cases. Subjects read the background information and the results of the five audit procedures (cues) of a case. Then, they made a judgment of the risk of misstatement for each case and clicked on the appropriate answer. Once the subject clicked the answer, whether it was correct or not, a message with the correct risk assessment was shown on the screen right below the answer scale. The message read: "The risk assessment of this case made by a very experienced partner, who is recognised as an expert in this audit task, is 60" (assuming the partner judged the risk of misstatement in this case was 60). Therefore, the subject had his/her own judgment together with the correct judgment of that case on one screen. Appendix B shows the presentation of OFB information on the computer screen. After reviewing the information, they proceeded to the next case by following the instructions on the screen (clicking the "NEXT" button shown on the screen).

TPF, CFB, and CombFB were presented to subjects belonging to the respective treatment groups before they started the second block of 16 trials. When they finished Stage 3, the researcher provided (in Excel files) the relevant feedback information. Subjects were allowed to study the information for as long as they liked. They could proceed to the second block of 16 trials (Stage 4) by following the instructions at the end of the feedback information (I.e. by clicking on an icon at the bottom of the screen).

4.5.1 Task Properties Feedback - Contents and Presentation

TPF refers to information about the relationship between cues (audit procedures) and the criterion (the correct risk judgment). In this study, TPF was information about the importance of each cue and its combination with other cues in influencing the risk assessment. Following Brown and Solomon (1990; 1991), the importance was defined as the influence of each cue on the amount of variation in the partner's risk assessments in the respective task.

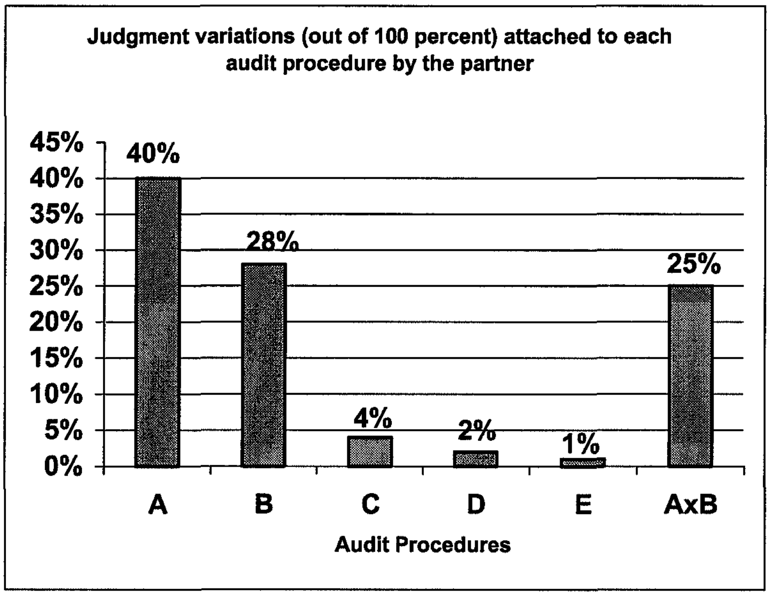
For Task 1B, the cues related to the risk assessments in a linear manner and there was no interaction effect among the cues. The relative importance of the cues was presented in the form of bar charts with the numerical percentage also shown on the chart. The presentation of feedback information for Task 1B was similar to that for Task 1A (see Exhibit 4.6), except that there was no interaction effect (that is, interaction of Procedures A and B). As there was no interaction effect, only the bar chart was presented. My presentation of feedback information followed the most common approach adopted in psychology studies. After extensive review of the psychology literature, Balzer et al. (1989) conclude that the major method of presenting TPF and CFB has been via graphics.

For Task 1A, the presentation of feedback information to subjects who completed this task is more complex. Two of the cues (Procedures A and B) are expected to relate to the risk assessment in a compensatory-form ordinal interaction (Brown and Solomon 1991). To make sure that subjects understood the feedback information, both bar and line charts were presented to them, together with a brief narrative description to explain the meaning of the interaction effect of these two cues. The TPF information presented to subjects who completed Task 1A is depicted in Exhibit 4.6.

Exhibit 4.6

Task Properties Feedback Information Provided to Subjects Completing Task 1A

Shown below are the amount of variations (or relative importance) in the judgments related to each of the five audit procedures by a very experienced partner, who is recognised as an expert in this audit task. In making your judgments on cases 17 to 32 (in Stage 4 of the study), please use the information provided by the partner about the importance of the five audit procedures in making your A/R risk assessments.



Procedure A is "Confirm a 95% reliability sample of year-end A/R (positive form)".
Procedure B is "Verify sales transactions for and observe collections of a 95% reliability sample of year-end A/R".
Procedure C is "Reconcile A/R subsidiary ledger to the general ledger control account".
Procedure D is "Review aged A/R listing for related party transactions, long-term notes, etc".
Procedure E is "Review Board of Directors meeting minutes and correspondence files to determine if receivables were factored or discounted".

Exhibit 4.6 (cont'd)

Interpretation of the bar chart:

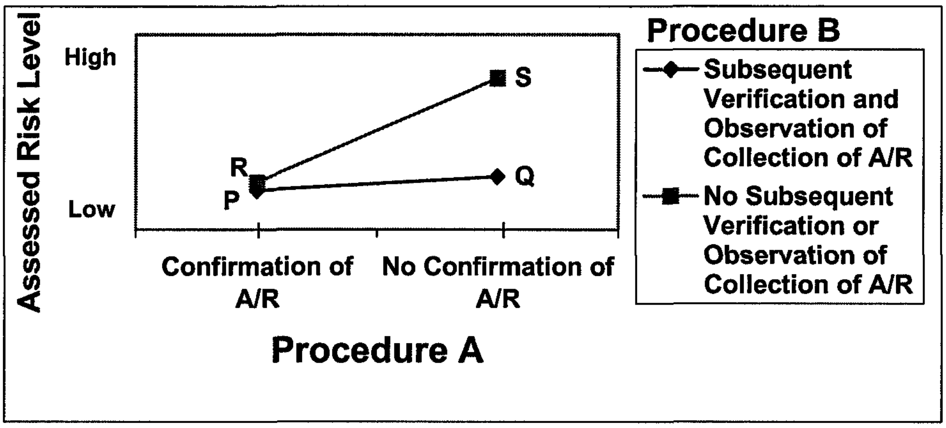
For the partner, Procedure A was the most important audit procedure, accounting for 40% of all the variations of the risk assessments that the year-end A/R balance could be materially misstated as a result of non-existence (the risk assessments).

Procedure B was the second most important, accounting for 28% of all the variations in the partner's A/R risk assessments.

Procedures C, D, and E were less important in the partner's risk assessments and respectively accounted for 4%, 2%, and 1% of all the variations in the A/R risk assessments.

Apart from the importance of Procedure A or Procedure B considered individually, these two procedures ***considered jointly*** accounted for another 25% of all the variations in the partner's A/R risk assessments.

This joint impact of these two audit procedures in the partner's risk assessments can be shown diagrammatically as follows:



Interpretation of the line chart:
You will notice that the lowest risk assessment occurs at P (both procedures have been completed), but the risk assessment does not increase much when only one of the procedures is completed (Q and R). However, when neither procedure has been completed (S), the risk assessment is considerably higher.

Please click the "Microsoft Visual FoxPro" icon at the lower left-hand corner of the screen to continue after you studied the above information. Thank you.

4.5.2 Cognitive Feedback - Contents and Presentation

CFB provides information about a subject's cognitive strategy. A commonly used CFB is the relationship between the cues and the subject's judgments (Remus et al. 1996). In this study, I modeled subjects' judgments based on the first block of 16 cases. As the 16 cases represented a one-half fractional replication of a 2^5 factorial design, only the main effects and 10 two-way interaction effects were estimated. Following Brown and Solomon (1991), the weight of each cue (audit procedure) was "computed by dividing the sum of the squares for the term by the total sum of squares for the model" (p. 106).

For each subject receiving CFB, their respective judgment weight for each of the 5 cues (the main effects) was included in the feedback information. For the 10 possible two-way interactions, I only presented a maximum of two interactions, if they satisfied the following conditions:

(i) the weight of each interaction was greater than, or equal to 4%;

(iii) if there were three or more interactions which satisfied the first condition, I only selected the two interactions which accounted for the greatest portion of the subject's judgment variation.

The decision to present a maximum of two interactions was made to avoid information overload for the subjects. This was particularly the case for subjects in the CombFB groups. Also, it would have been difficult for subjects to interpret many interaction terms. The feedback information in the CFB was very similar to that of TPF with the exception that the description and explanation clearly indicated they were the judgment weights of the subject.

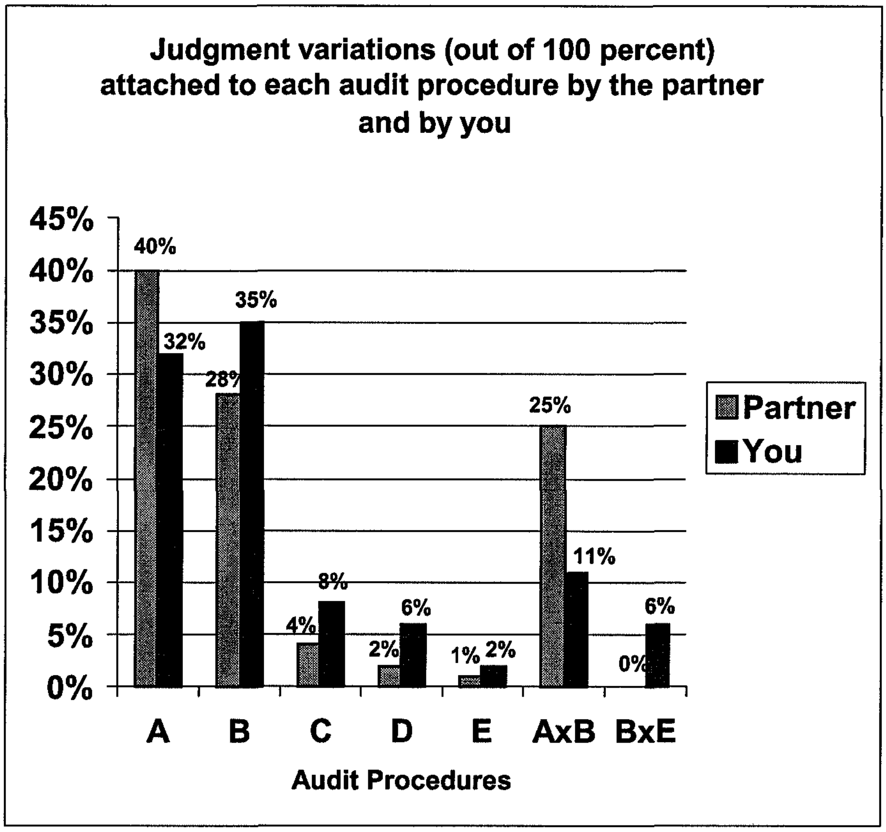
4.5.3 Combined Feedback - Contents and Presentation

Subjects in the CombFB groups were provided with both TPF and CFB (combined feedback). This has been the common practice adopted in previous studies investigating CombFB (see for example, Remus et al. 1996). A sample of the CombFB information presented to subjects who completed Task 1A is shown in Exhibit 4.7.

Exhibit 4.7

Combined Feedback Information Provided to Subjects Completing Task 1A

Shown below are the amount of variations (or relative importance) in the judgments related to each of the five audit procedures by you and by a very experienced partner, who is recognised as an expert in this audit task. In making your judgments on cases 17 to 32 (in Stage 4 of the study), please use the information provided by the partner about the importance of the five audit procedures and your own importance weights in making your A/R risk assessments.



Procedure A is "Confirm a 95% reliability sample of year-end A/R (positive form)".
Procedure B is "Verify sales transactions for and observe collections of a 95% reliability sample of year-end A/R".
Procedure C is "Reconcile A/R subsidiary ledger to the general ledger control account".
Procedure D is "Review aged A/R listing for related party transactions, long-term notes, etc".
Procedure E is "Review Board of Directors meeting minutes and correspondence files to determine if receivables were factored or discounted".

Exhibit 4.7 (continued)

Interpretation of the bar chart:

For the partner, Procedure A was the most important audit procedure, accounting for 40% of all the variations of risk assessments that the year-end A/R balance could be materially misstated as a result of non-existence (the risk assessments).

For you, Procedure B was the most important and accounted for 35% of all your variations of your risk assessments.

For the partner, Procedure B was the second most important, accounting for 28% of all the variations in the partner's A/R risk assessments.

For you, Procedure A was the second most important, accounting for 32% of all your variations of your risk assessments.

For the partner, Procedures C, D, and E were less important and respectively accounted for 4%, 2%, and 1% of all the variations of the partner's A/R risk assessments.

For you, Procedures C, D, and E were less important and respectively accounted for 8%, 6%, and 2% of all the variations of your A/R risk assessments.

Apart from the importance of Procedure A or Procedure B considered individually, these two procedures ***considered jointly*** accounted for another 25% of the partner's (11% of your) variations in the A/R risk assessments.

For you, apart from the importance of Procedure B and Procedure E considered individually, these two procedures ***considered jointly*** accounted for another 6% of all your variations in your A/R risk assessments.

This joint effect of these two audit procedures for the partner can be shown diagrammatically as follows:

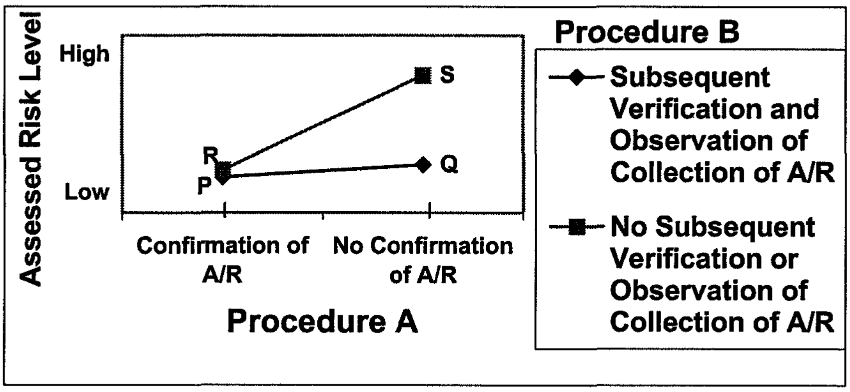
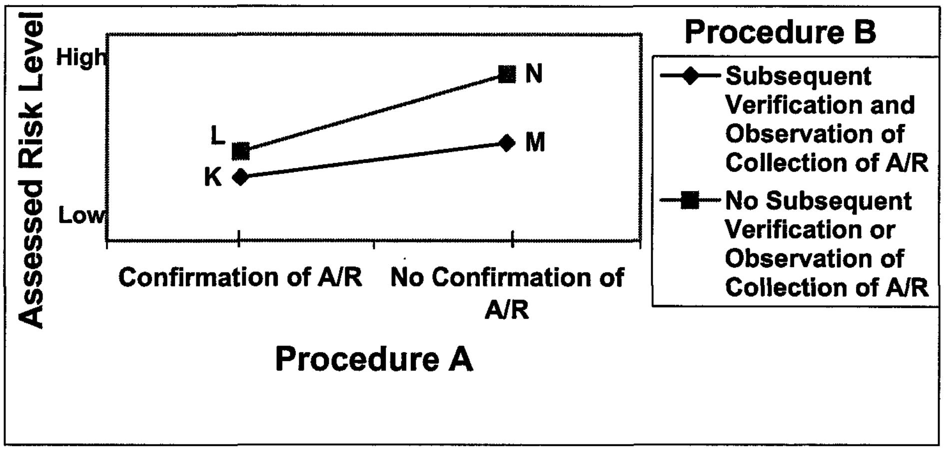


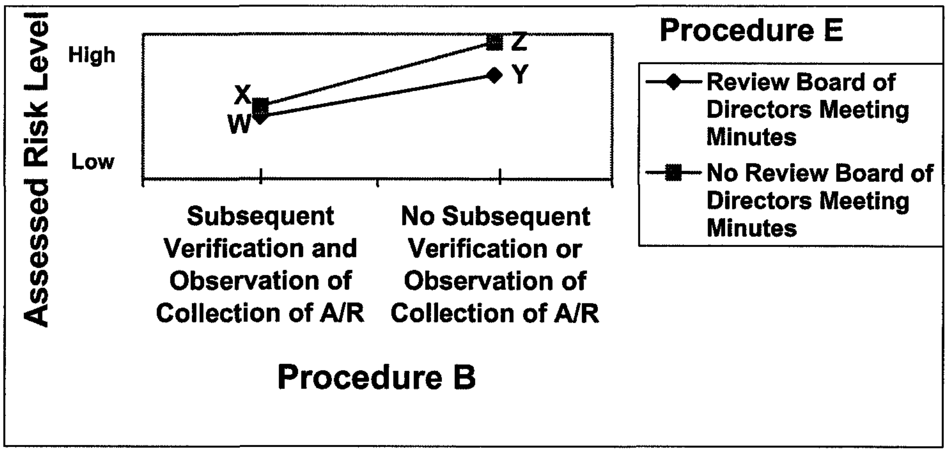
Exhibit 4.7 (continued)

Interpretation of the line chart:

You will notice that the lowest risk assessment occurs at P (both procedures have been completed), but the risk assessment does not increase much when only one of the procedures is completed (Q and R). However, when neither procedure has been completed (S), the risk assessment is considerably higher.

The two joint effects of two audit procedures for YOU can be shown diagrammatically as follows:





The interpretation of the above charts is similar to that of the partner.

Please click the "Microsoft Visual FoxPro" icon at the lower left-hand corner of the screen to continue after you studied the above information. Thank you.

4.6 Measurement of Variables

4.6.1 Judgment Performance

Judgment performance was measured using two approaches. First, judgment accuracy was measured. Subjects' risk judgments were compared to the selected audit partner's judgments (discussed in section 4.5) and the mean absolute error or difference (MAE) provided the measure of subjects' judgement accuracy. The smaller the MAE, the more accurate was the subject's judgment.

The second approach measured judgment achievement (r_a), a lens model parameter, which is a commonly used measure of judgment performance. Judgment achievement was computed for each subject by correlating his/her actual risk assessments with the assessments of the selected partner described in Section 4.5. This measure of judgment performance was included because of its widespread use in previous studies, thus facilitating comparison, and allowing analysis of other lens model statistics.

4.6.2 Self-insight

Subjects were asked to externalise their judgment policies by rating the importance of each of the five cues in their risk assessments. The rating used a 100-point scale where 0 was not important at all and 100 was very important. Rating of cue importance has been found to be better than point allocation in terms of being (1) closer to theoretical valid weight (Zhu and Anderson 1991), (2) easier for subjects to perform (Doyle et al 1997), and (3) of high test-retest reliability (Bottomley et al. 2000).

For each subject, the relative weight assigned to each audit procedure (cue) was calculated by dividing the points assigned to that cue by the sum of the points awarded to all five cues. That is, the sum of the weights of all five cues should equal one. These are referred to as the subjective weights.

In addition, an ANOVA model was computed for each subject based on the first block of 16 trials (which represents a one-half fractional replication of a 2^5 factorial design). Each subject's model estimated all five main effects and ten two-way interactions. However, higher order interactions could not be estimated and are assumed to be negligible. The variation in judgment attributed to each audit procedure (cue) in a subject's model was computed by dividing the sum of the squares for each cue by the total sum of squares for that model. These are referred to as the objective weights.

An index of self-insight for each subject was computed by correlating a subject's subjective weights with their objective weights. I classified subjects as either having high or low level of self-insight by comparing a subject's self-insight index with the respective median for the two tasks. Subjects who completed Task 1A (Task 1B) were classified by comparing their response with the median of subjects in groups 1 to 5 (groups 6 to 10).

4.6.3 Configural Information Processing

Subjects' use of configural information processing was assessed by the explained judgment variance in the predicted interaction term in each subjects' ANOVA model. The judgment variance attributed to each cue within an auditor-model was computed by dividing the sum of the squares for that cue by the total sum of the squares for the model. Following Brown and Solomon (1990; 1991), a subject was considered as having applied configural cue processing if the judgment model of that subject attributed more than 4% of total judgment variance to the cues expected to have an interaction effect based on domain- and task-specific knowledge. The threshold of 4% was used for two reasons.

(1) This threshold was used by Brown and Solomon (1991) after extensive testing. "Pilot study (n=16) results in which a full (non-fractional) 2^5 factorial design and similar, although not identical, experiment materials and procedures were used indicated that no three-, four- or five-way interactions exceeded four percent of total judgment variance" (p. 29).

(2) A Monte Carlo simulation was conducted with a 2^5 full factorial model of 32 cases over 100,000 trials. I generated random responses for each set of 32 cases and calculated the weight of (or judgment variance attributed to) each of the main and interaction effects. I found that 95% of cue weights for the two-way interaction terms were equal to or less than 3.9%, which was very close to the 4% threshold used in previous studies. This provided evidence to support the use of the 4% threshold used in previous studies. This threshold was also adopted in this study in evaluating whether

different feedback conditions lead more subjects to exhibit configural cue processing.

4.6.4 Lens Model Parameters

The lens model parameters for each subject were derived from the responses of each subject and that of the benchmark (as described above). The judgment achievement index (r_a) was computed by correlating the responses of the subjects with those of the relevant partner. The consistency index (Rs) was computed by correlating the subject's judgments with those generated from his/her judgment model. Subject's knowledge of the linear component of the task (G) was computed by correlating the judgments of the relevant partner's model with those of the subject's model. Subjects' knowledge of the non-linear component of the task was computed by correlating the residual of the relevant partner's judgments (the difference between the partner's actual judgments and that of their model) and the residual of the subject's judgments. All the lens model parameters were transformed using the Fisher z-transformation for subsequent data analysis.

4.7 Statistical Analysis

Judgment performance (in terms of MAE and r_a), judgment consistency, and task knowledge were analysed separately via repeated measures ANOVAs involving a 5 (feedback treatments) x 2 (task complexity) x 2 (blocks) design. Whenever there was a significant difference found in the ANOVA model, multiple comparisons (using Tukey HSD tests) or t-tests were conducted, to identify the source of the difference. In assessing the number of subjects exhibiting above-criterion configural cue processing, after different feedback treatments, Kruskal-Wallis and chi-square tests were conducted. Non-parametric tests were adopted because a subject may or may not exhibit above-criterion configural cue processing, thus providing nominal data.

CHAPTER 5

RESULTS

After eliminating subjects who failed the manipulation checks (see below), there were 224 Big-5 auditors in Hong Kong who participated in the study. Table 5.1 presents the distribution of subjects across the treatment groups in the experiment. The distribution of subjects across treatment groups is quite even. There is no statistically significant difference between the order of cue presentation and judgment performance in each treatment group. Therefore, the subjects in both cue presentation orders in each treatment group are combined together in further data analysis.

5.1 Manipulation Checks

Subjects in the TPF, CFB, and CombFB groups were presented with the appropriate feedback before they started the second block of trials. After they had completed the second block of trials, they were asked to recall and rank the three most important audit procedures from a list of all five audit procedures (cues). There were seven subjects (3 in the TPF groups, 1 in the CFB groups, and 3 in the combined feedback groups) who reported a very different ranking of importance of the audit procedures. Six of them did not rank the two most important procedures in the correct order. One of them did not include the most important procedure in the three choices. Therefore, their responses were excluded.

Table 5.1
Distribution of Subjects in Each Treatment Group

Treatment Group No.	Task	Order of Cues	Feedback Provided	No. of Subjects
1.1	1A	1	NFB	11
1.2	1A	2	NFB	11
2.1	1A	1	OFB	11
2.2	1A	2	OFB	12
3.1	1A	1	TPF	11
3.2	1A	2	TPF	10
4.1	1A	1	CFB	11
4.2	1A	2	CFB	12
5.1	1A	1	CombFB	11
5.2	1A	2	CombFB	11
6.1	1B	1	NFB	12
6.2	1B	2	NFB	10
7.1	1B	1	OFB	11
7.2	1B	2	OFB	11
8.1	1B	1	TPF	11
8.2	1B	2	TPF	13
9.1	1B	1	CFB	11
9.2	1B	2	CFB	12
10.1	1B	1	CombFB	11
10.2	1B	2	CombFB	11
Total				224

Note: Task 1A involves configural cues
Task 1B does not involve configural cues
Order 1 where cues A and B are listed first (A, B, C, D, E)
Order 2 where cues A and B are listed last (D, E, C, A, B)
NFB = subjects receive no feedback during the experiment (control group)
OFB = subjects receive outcome feedback after each case in the first block of cases
TPF = subjects receive task properties feedback after the first block of cases
CFB= subjects receive cognitive feedback after the first block of cases
CombFB = subjects receive both TPF and CFB after the first block of cases

A second manipulation check compares the task complexity of the two cases. It is suggested that Task 1A involves configural cues and demands more cognitive resources and intermediate steps in making judgments on the part of subjects. It is high in task complexity when compared with Task 1B, which does not involve configural cue processing to complete. It is expected that subjects completing Task 1A would require more time than their counterparts completing Task 1B. Independent samples t-tests results confirm this prediction. In the first block of cases, subjects who completed Task 1A and Task 1B took an average of 12.19 and 11.25 minutes, respectively. The difference in means is significant ($t_{221} = 4.13$, $p < 0.001$). In completing the second block of cases the average time taken to complete Task 1A and Task 1B was 10.03 and 9.10 minutes respectively. The difference in average time taken was also significant ($t_{221} = 4.44$, $p < 0.001$).

The results indicate that subjects on average took more time to complete Task 1A, which involves configural information processing in making the risk assessments as specified by domain- and task-specific knowledge. Thus, it involves a higher level of task complexity than Task 1B. Post-experiment discussion with subjects also supported this.

5.2 Statistical Analysis

Judgment performance was analysed with a repeated measures analysis of variance involving a 5 (feedback treatments) x 2 (task complexity levels) x 2 (blocks) design. Feedback and task complexity manipulated between-subjects and block was a within-subjects variable. Judgment performance was measured using judgment accuracy (MAE) and judgment achievement (r_a). The lower (higher) the MAE (r_a), the more accurate was the subject's judgment. That is, the subject's judgments correspond more closely to that of the expert audit partner. The descriptive statistics of judgment accuracy (MAE) and judgment achievement (r_a) are presented in Table 5.2.

The results of the overall ANOVA are presented in Table 5.3. The main effects for both the between-subjects variables (Task and Feedback) were significant, indicating that subjects' judgment accuracy was significantly different between the two tasks. The significant Task main effect was due to the fact that judgment accuracy in Task 1A was significantly lower than that in Task 1B (consistent with Task 1A being a more complex task). Also, the various feedback conditions had significantly different effects on subjects' judgment accuracy. Multiple comparisons (using Tukey HSD tests) of the different feedback groups, with respect to average judgment over the two blocks of trials, indicated significant differences between the various feedback groups which are discussed below. This is consistent with the significant Feedback main effect. However, the Task * Feedback interaction effect was not significant. The various feedback conditions did not have statistically different effects across the tasks.

Block was the within-subjects variable and was significant in the ANOVA model. This suggested that there was a significant difference in the judgment accuracy (MAE) across the two blocks of trials.

Table 5.2
Descriptive Statistics of Judgment Accuracy (MAE) and Judgment Achievement (r_a)

Panel A: Group average MAE and average r_a in Task 1A (configural case)
(standard deviation shown in paretheses)

	Block 1						Block 2					
	NFB	OFB	TPF	CFB	CombFB	Overall	NFB	OFB	TPF	CFB	CombFB	Overall
MAE	17.67 (4.42)	16.39 (4.33)	16.61 (6.22)	18.37 (5.77)	17.64 (5.44)	17.34 (5.24)	17.61 (5.34)	15.95 (4.17)	12.68 (4.09)	14.08 (3.57)	12.24 (2.65)	14.54 (4.46)
ra	0.55 (0.20)	0.61 (0.18)	0.55 (0.29)	0.51 (0.26)	0.55 (0.21)	0.55 (0.23)	0.55 (0.21)	0.62 (0.17)	0.73 (0.12)	0.70 (0.12)	0.75 (0.06)	0.67 (0.16)

Panel B: Group average MAE and average r_a in Task 1B (non-configural case)
(standard deviation shown in paretheses)

	Block 1						Block 2					
	NFB	OFB	TPF	CFB	CombFB	Overall	NFB	OFB	TPF	CFB	CombFB	Overall
MAE	16.42 (5.20)	15.26 (4.18)	15.44 (3.69)	15.90 (3.78)	16.48 (4.33)	15.89 (4.21)	16.62 (5.44)	11.39 (3.48)	11.35 (3.54)	14.81 (3.92)	12.67 (4.71)	13.35 (4.67)
r_a	0.60 (0.21)	0.63 (0.17)	0.65 (0.13)	0.63 (0.15)	0.61 (0.15)	0.63 (0.16)	0.60 (0.21)	0.77 (0.09)	0.78 (0.10)	0.67 (0.15)	0.74 (0.15)	0.71 (0.16)

Table 5.3

Overall ANOVA Testing for Judgment Accuracy (MAE)

	SS	df	MS	F ratio	P
Between subjects:					
Task	186.15	1	186.15	6.082	0.014
Feedback	500.84	4	125.21	4.091	0.003
Task * Feedback	77.73	4	19.43	0.635	0.638
Error	6,549.53	214	30.61		
Within subjects:					
Block	800.92	1	800.92	82.55	0.000
Block * Task	2.40	1	2.40	0.25	0.620
Block * Feedback	293.67	4	73.42	7.57	0.000
Block * Task * Feedback	137.11	4	34.28	3.53	0.008
Error	2,076.35	214	9.70		

The three-way interaction (Block * Task * Feedback) was significant. Tables 5.4 and 5.5 show the mean group judgment accuracy across blocks and task complexity. The mean group judgment accuracy in Table 5.4 suggests that the significant three-way interaction was mainly due to a strong Block * Feedback interaction. For example, TPF, CFB, and CombFB led to significant judgment accuracy improvement in block 2 in Task 1A. However, OFB and NFB groups did not significantly improve their performance. For the low complexity task (Task 1B), only OFB, TPF, and CombFB led to significant improvement in judgment accuracy in block 2, but this was not the case for NFB and CFB. Mean group judgment accuracy (see Tables 5.4 and 5.5) also indicated that Block * Task interaction was not significant. This suggested that the difference in improvement in judgment accuracy (MAE) between two blocks of trials was not significantly different across the two tasks. For Task 1A and Task 1B, the average improvement in judgment accuracy (MAE) in block 2 over block 1 was 2.80 and 2.54 respectively (see Tables 5.4 and 5.5).

The pair-wise comparisons in Panel B of Tables 5.4 and 5.5 show that there was a significant difference in judgment improvement across the two blocks of trials for the different feedback conditions. Different feedback conditions had different effects in improving judgment performance in tasks of varying levels of complexity. In a task of high complexity involving configural cues (Task 1A), subjects in the TPF, CFB, and CombFB groups had a significantly higher improvement in judgment accuracy (MAE) than their NFB and OFB counterparts. In a task of low complexity involving non-configural cues (Task 1B), subjects in the OFB, TPF, and CombFB groups showed significantly higher improvements in judgment accuracy than subjects in the NFB and CFB groups.

The TPF and CombFB groups were more effective in improving subjects' judgment accuracy than the NFB groups in tasks of both low and high complexity. OFB was more effective in improving judgment accuracy than the NFB group in a low complexity task only. On the other hand, CFB was effective in a task of high complexity (involving configural cues), but not in a low complexity one.

Table 5.4
Judgment Accuracy (MAE) for Each Feedback Treatment in Task 1A

Panel A: Average Judgment Accuracy (MAE) in Both Blocks of Trials in Task 1A (with configural cues)

Feedback	n	Mean MAE # in 1st Block	Mean MAE in 2nd Block	Difference in Mean MAE
NFB	22	17.67	17.61	0.06
OFB	23	16.39	15.95	0.44
TPF	21	16.61	12.68	3.93***
CFB	23	18.37	14.08	4.29***
CombFB	22	17.64	12.24	5.40***
Average		17.34	14.54	2.80

no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

Panel B: Pairwise Comparison Using the Tukey HSD test on Mean Difference in MAE between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	0.38	3.87**	4.23***	5.34***
OFB		0	3.49**	3.85**	4.96***
TPF			0	0.36	1.47
CFB				0	1.11
CombFB					0

** sig. at 0.05 level
*** sig. at 0.01 level

Table 5.5
Judgment Accuracy (MAE) for Each Feedback Treatment in Task 1B

Panel A: Average Judgment Accuracy (MAE) in Both Blocks of Trials for Task 1B (with non-configural cues)

Feedback	n	Mean MAE # in 1st Block	Mean MAE in 2nd Block	Difference in Mean MAE
NFB	22	16.42	16.62	-0.20
OFB	22	15.26	11.39	3.87***
TPF	24	15.44	11.35	4.09***
CFB	23	15.90	14.81	1.09
CombFB	22	16.48	12.67	3.81***
Average		15.89	13.35	2.54

no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

Panel B: Pairwise Comparison Using the Tukey HSD test on Difference in MAE between the First and the Second Blocks of Trials

	Difference in Mean (Column - Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	4.07**	4.29**	1.29	4.01**
OFB		0	0.22	-2.78	-0.06
TPF			0	-3.00	-0.28
CFB				0	2.72
CombFB					0

** sig. at 0.05 level
*** sig. at 0.01 level

Judgment achievement (r_a), a lens model parameter, is another commonly used criterion for judgment performance in both the accounting and psychology literatures. The repeated measures ANOVA results using r_a as the dependent variable, and the judgment achievement across two blocks of trials for Task 1A and Task 1B as the independent variable are presented in Tables 5.6, 5.7, and 5.7. The results are very similar to those using MAE as a criterion to test judgment performance. The two judgment performance criteria (MAE and r_a) were highly correlated and the correlation was significant. The correlation and significance in the first and second blocks was –0.959 ($p = 0.000$) and –0.956 ($p = 0.000$), respectively. The negative correlation was due to the fact that a high MAE (r_a) indicated a lower (higher) judgment performance. The results were in line with that of A. Ashton (1985), who found in two accounting decision making tasks that judgment achievement or consensus (r_a) implied accuracy (MAE).

Table 5.6

Overall ANOVA Testing for Judgment Achievement (r_a)

	SS	df	MS	F ratio	P
Between subjects:					
Task	1.029	1	1.029	9.224	0.003
Feedback	1.814	4	0.453	4.064	0.003
Task * Feedback	0.222	4	0.056	0.498	0.737
Error	23.879	214	0.112		
Within subjects:					
Block	3.546	1	3.546	78.819	0.000
Block * Task	0.000	1	0.000	0.007	0.933
Block * Feedback	1.227	4	0.307	6.821	0.000
Block * Task * Feedback	0.594	4	0.149	3.301	0.012
Error	9.626	214	0.045		

Table 5.7

Judgment Achievement (r_a) for Each Feedback Treatment in Task 1A

Panel A: Average Judgment Achievement (r_a) in Both Blocks of Trials in Task 1A (with configural cues)

Feedback	n	Mean r_a # in 1st Block	Mean ra in 2nd Block	Difference in Mean ra
NFB	22	0.5520	0.5498	-0.0022
OFB	23	0.6069	0.6196	0.0127
TPF	21	0.5520	0.7295	0.1775***
CFB	23	0.5091	0.7006	0.1915***
CombFB	22	0.5496	0.7512	0.2016***
Average		0.5540	0.6694	0.1154

#1 no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

#2 higher r_a indicated higher correlation between subjects' judgments and the benchmark and was higher in judgment achievement or consensus

Panel B: Pairwise Comparison Using Tukey HSD test on Mean Difference in r_a between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	0.0149	0.1797**	0.1937**	0.2038***
OFB		0	0.1648**	0.1788**	0.1889***
TPF			0	0.0014	0.0241
CFB				0	0.0101
CombFB					0

** sig. at 0.05 level
*** sig. at 0.01 level

Table 5.8

Judgment Achievement (r_a) for Each Feedback Treatment in Task 1B

Panel A: Average Judgment Achievement (r_a) in Both Blocks of Trials in Task 1B (with non-configural cues)

Feedback	n	Mean r_a # in 1st Block	Mean ra in 2nd Block	Difference in Mean ra
NFB	22	0.6007	0.5986	-0.0021
OFB	22	0.6344	0.7740	0.1396***
TPF	24	0.6525	0.7782	0.1257***
CFB	23	0.6301	0.6734	0.0433
CombFB	22	0.6085	0.7382	0.1297***
Average		0.6258	0.7133	0.0875

#1 no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

#2 higher r_a indicated higher correlation between subjects' judgments and the benchmark and was higher in judgment achievement or consensus

Panel B: Pairwise Comparison Using Tukey HSD test on Mean Difference in r_a between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	0.1417**	0.1278**	0.0545	0.1318**
OFB		0	-0.0139	-0.0963	-0.0099
TPF			0	-0.0824	0.0040
CFB				0	0.0864
CombFB					0

** sig. at 0.05 level

*** sig. at 0.01 level

5.3 Testing of Hypotheses - Feedback

5.3.1 Hypothesis H1a

H1a hypothesises that OFB is more effective in improving judgment performance in a non-configural task (a task of low complexity, Task 1B) than a configural task (a task of high complexity, Task 1A). In testing this hypothesis, an ANOVA was run with the judgment accuracy (MAE) of subjects in the OFB groups as the dependent variable and task complexity and block as the independent variables. Again task complexity is a between-subjects factor and block is a within-subjects factor. A significant Block * Task Complexity effect in the expected direction would indicate that OFB has differential effectiveness in improving judgment accuracy in tasks of different levels of complexity. The results of the ANOVA are presented in Table 5.9. The Block * Task interaction was significant ($F_{(1,43)} = 7.057$, $p = 0.011$) and the difference was in the expected direction (see Panel B of Table 5.9). This interaction effect is shown in Figure 5.1. OFB was effective in improving judgment accuracy in a task of low complexity (Task 1B), but not in a task of high complexity, which involved configural cues (Task 1A).

The significant interaction is consistent with my earlier results showing that the improvement in judgment accuracy (MAE) for Task 1A (0.44) was not significant (see Table 5.4) whilst that of Task 1B (3.87) was significant (see Table 5.5). The significant main effect for task complexity was again due to the higher performance on the non-configural case on average. Panel B of Table 5.9 also indicates that the significant Block main effect was mainly due to the significant improvement in judgment accuracy in Task 1B.

Table 5.9

Results of Judgment Accuracy (MAE) for Two OFB Groups across Two Tasks

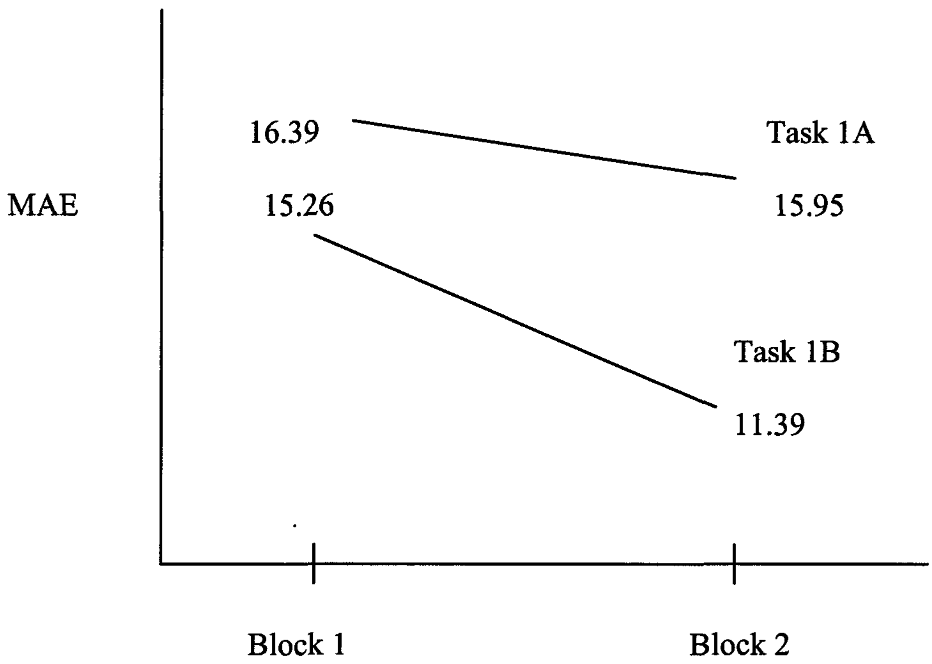
Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	181.98	1	181.98	7.723	0.008
Error	1,013.19	43	23.56		
Within subjects:					
Block	103.88	1	103.88	11.090	0.002
Block * Task	66.10	1	66.10	7.057	0.011
Error	402.76	43	9.37		

Panel B: Average Judgment Accuracy (MAE) in the Two Blocks of Trials

	Task 1A	Task 1B
Block 1	16.39	15.26
Block 2	15.95	11.39
Difference	0.44	3.87

Figure 5.1: Average Judgment Accuracy (MAE) for Two Blocks of Trials across Two Tasks



The results of an ANOVA with judgment achievement (r_a) as the dependent variable also showed a similar pattern. The results are presented in Table 5.10. These results are again consistent with earlier analysis in Tables 5.7 and 5.8.

Overall, both the results of judgment accuracy (MAE) and judgment achievement (r_a) support H1a. OFB was more effective in improving judgment performance in a task of low complexity (without configural cues) than in a task of high complexity (with configural cues).

Table 5.10

ANOVA Results for Judgment Achievement (r_a) for Two OFB Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	0.655	1	0.655	6.528	0.014
Error	4.315	43	0.100		
Within subjects:					
Block	0.469	1	0.469	10.630	0.002
Block * Task	0.350	1	0.350	7.939	0.007
Error	1.897	43	0.044		

Panel B: Average Judgment Achievement (r_a) in Two Blocks of Trials

	Task 1A	Task 1B
Block 1	0.6069	0.6344
Block 2	0.6196	0.7740
Difference	0.0127	0.1396

5.3.2 Hypothesis H1b

H1b predicts that TPF is effective in improving judgment performance in tasks involving both configural and non-configural processing. A repeated measures ANOVA was run with the judgment accuracy (MAE) of subjects in the two TPF groups in the two blocks of trials as the dependent variable and task complexity as the independent variable with the block as a within-subjects factor. To support the hypothesis, there should be a significant Block main effect and a non-significant

Block * Task interaction effect. The ANOVA and other results are presented in Table 5.11.

Table 5.11

Results of Judgment Accuracy (MAE) Improvement in Two TPF Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	34.69	1	34.69	1.025	0.317
Error	1,454.53	43	33.83		
Within subjects:					
Block	359.94	1	359.94	60.204	0.000
Block * Task	0.14	1	0.14	0.024	0.878
Error	257.08	43	5.98		

Panel B: Average Judgment Accuracy (MAE) in Two Blocks of Trials

	Task 1A	Task 1B
Block 1	16.61	15.44
Block 2	12.68	11.35
Difference	3.93	4.09

There was a highly significant Block main effect. The MAE figures in Panel B of Table 5.11 suggest that TPF was effective in improving judgment accuracy in both tasks between two blocks of trials. TPF led to an average improvement in MAE of 3.93 and 4.09 for Task 1A and Task 1B, respectively. There was no significant difference in judgment accuracy between the two tasks. This was supported by the non-significant Block * Task interaction effect ($F_{(1,43)} = 0.024$, $p = 0.878$). This result was consistent with those shown in Tables 5.4 and 5.5. It was shown there that TPF led to a significant improvement in judgement accuracy in the second block of trials when considered individually for each task.

The ANOVA and Tukey HSD test results using judgment achievement (r_a) as the dependent variable produced similar findings, which are presented in Table 5.12. These results are consistent with earlier results shown in Tables 5.7 and 5.8. On the whole, the results support H1b.

Table 5.12

Results of Judgment Achievement (r_a) Improvement in Two TPF Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	0.329	1	0.329	2.646	0.111
Error	5.339	43	0.124		
Within subjects:					
Block	1.673	1	1.673	51.831	0.000
Block * Task	0.000	1	0.000	0.019	0.890
Error	1.388	43	0.032		

Panel B: Average Judgment Achievement (r_a) in Two Blocks of Trials

	Task 1A	Task 1B
Block 1	0.5520	0.6525
Block 2	0.7295	0.7782
Difference	0.1775	0.1257

5.3.3 Hypothesis H1c

H1c predicts that CFB is more effective in improving judgment performance in a task involving configural cue processing (high complexity) than in one that does not (low complexity). To test H1c, a repeated measures ANOVA was conducted. It was similar to that conducted when testing hypothesis H1b, except that judgment accuracy (MAE) in the two CFB groups was the dependent variable. If CFB had a differential effect in improving judgment achievement between the two tasks, there would have been a significant Block * Task interaction effect. The results are presented in Table 5.13.

The Block main effect was highly significant ($F_{(1,44)} = 26.576$, $p = 0.000$) indicating that subjects' judgment achievement improved between the two blocks of trials. Panel B of Table 5.13 suggests that the Block main effect was largely attributed to the improvement in Task 1A. This is consistent with the earlier analysis that CFB led to a significant (non-significant) improvement in judgment achievement in block 2 in Task 1A (Task 1B) as shown in Panel A of Table 5.4 (Table 5.5). The Block * Task interaction was also significant ($F_{(1,44)} = 9.439$, $p = 0.004$). This interaction effect is depicted in Figure 5.2.

Table 5.13

Results of Judgment Accuracy (MAE) Improvement in Two CFB Groups across Two Tasks

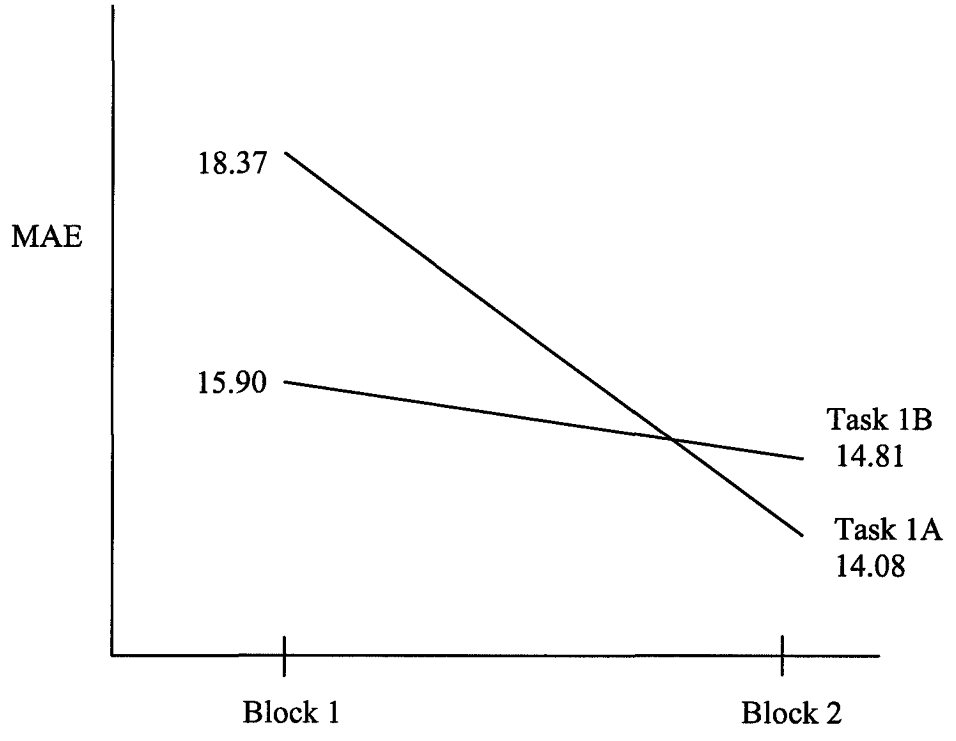
Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	17.39	1	17.39	0.550	0.462
Error	1,391.49	44	31.63		
Within subjects:					
Block	166.46	1	166.46	26.576	0.000
Block * Task	59.12	1	59.12	9.439	0.004
Error	275.59	44	6.26		

Panel B: Average Judgment Accuracy (MAE) in Two Blocks

	Task 1A	Task 1B
Block 1	18.37	15.90
Block 2	14.08	14.81
Difference	4.29	1.09

Figure 5.2: Average Judgment Accuracy (MAE) Improvement for CFB across Two Tasks



The ANOVA and Tukey HSD test results using judgment achievement (r_a) as the dependent variable were very similar to those shown above, and are presented in Table 5.14. They also support earlier comparisons shown in Tables 5.7 and 5.8. Based on the results of the two ANOVA models, hypothesis H1c is supported.

Table 5.14

Results of Judgment Achievement (r_a) for Two CFB Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	0.073	1	0.073	0.692	0.410
Error	4.664	44	0.106		
Within subjects:					
Block	0.713	1	0.713	18.653	0.000
Block * task	0.234	1	0.234	6.122	0.017
Error	1.681	44	0.038		

Panel B: Average Judgment Achievement (r_a) in Two Blocks of Trials

	Task 1A	Task 1B
Block 1	0.5091	0.6301
Block 2	0.7006	0.6734
Difference	0.1915	0.0433

5.3.4 Hypothesis H1d

H1d predicts that CombFB leads to greater improvement in judgment performance on a configural task than a non-configural task. A repeated measures ANOVA was run to test this hypothesis. Judgment accuracy (MAE) of subjects in the two blocks of trials of the CombFB groups in the two tasks was the dependent variable and the two independent variables were task complexity (a between-subjects factor) and block as a within-subjects factor. The results are depicted in Table 5.15.

Table 5.15
Results of Judgment Accuracy (MAE) Improvement in Two CombFB Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	3.00	1	3.00	0.123	0.728
Error	1,024.68	42	24.40		
Within subjects:					
Block	465.98	1	465.98	32.335	0.000
Block * Task	13.92	1	13.92	0.966	0.331
Error	605.26	42	14.41		

Panel B: Average Judgment Accuracy (MAE) in Two Blocks of Trials

	Task 1A	Task 1B
Block 1	17.64	16.48
Block 2	12.24	12.67
Difference	5.40	3.81

To support this hypothesis, the Block * Task interaction effect should be significant. The ANOVA results presented in Table 5.15 indicated that the Block * Task interaction effect was not significant ($F_{(1,42)} = 0.966$, $p = 0.331$). The judgment accuracy improvement was 5.40 and 3.81 for Task 1A and Task 1B, respectively. CombFB did not have a differential effect in improving judgment achievement in the two tasks. A similar result was found when judgment achievement (r_a) was the dependent variable in the ANOVA analysis (see Table 5.16). The result did not support H1d.

Table 5.16
Results of Judgment Achievement (r_a) Improvement in Two CombFB Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	0.064	1	0.064	0.785	0.381
Error	3.400	42	0.081		
Within subjects:					
Block	1.926	1	1.926	29.480	0.000
Block * task	0.010	1	0.010	0.146	0.704
Error	2.745	42	0.065		

Panel B: Average Judgment Achievement (r_a) in Two Blocks of Trials

	Task 1A	Task 1B
Block 1	0.5496	0.6085
Block 2	0.7512	0.7382
Difference	0.2016	0.1297

5.3.5 Hypothesis H1e

H1e predicts that the improvement in judgment performance between the CombFB and TPF groups is greater for a configural processing task than a non-configural processing task. To test this hypothesis, a repeated measures ANOVA was run with the judgment accuracy of subjects in the CombFB and TPF groups in the two blocks of trials as the dependent variable and task complexity as the independent variable with block as a within-subjects factor. To support this hypothesis, there should be a significant Block * Feedback * Task 3-way interaction. The results of the ANOVA are presented in Table 5.17. The only significant effect was the Block main effect. The Significant Block main effect indicated that both TPF and CombFB led to significant improvement in performance in the second block of trials in both tasks. All 2-way and 3-way interactions were not significant. The results suggested that TPF and CombFB did not lead to a significant differential effect in improving judgment accuracy across two tasks even though one of the tasks had configural processing and the other task did not.

Table 5.17
Results of Judgment Accuracy (MAE) for the TPF and CombFB Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Feedback	24.18	1	24.18	0.829	0.365
Task	28.90	1	28.90	0.991	0.322
Feedback*Task	8.50	1	8.50	0.291	0.591
Error	2,479.21	85	29.17		
Within subjects:					
Block	822.96	1	822.96	81.119	0.000
Block *Feedback	3.91	1	3.91	0.386	0.536
Block*Task	5.68	1	5.68	0.560	0.456
Block*Feedback*Task	8.51	1	8.51	0.838	0.362
Error	862.34	85	10.15		

Panel B: Average Judgment Accuracy (MAE) of TPF and CombFB Groups of Two Tasks in Two Blocks

Task 1A

	TPF (n = 21)	**CombFB (n = 22)**
Block 1	16.61	17.64
Block 2	12.68	12.24
Difference	3.93	5.40

Task 1B

	TPF (n = 24)	**CombFB (n = 22)**
Block 1	15.44	16.48
Block 2	11.35	12.67
Difference	4.09	3.81

A repeated measures ANOVA was run with judgment achievement (r_a) as the dependent variable with the same independent variables as above. The results are presented in Table 5.18, and they were similar to those discussed above, except that the Task main effect was marginally significant ($F_{(1,85)} = 3.301$, $p - 0.073$). There was a marginally significant difference in judgment achievement between the two tasks. Panel B of Table 5.18 shows the improvement in r_a in the two tasks. TPF and CombFB led to slightly higher improvement in r_a in a task involving configural processing (Task 1A) than in a task involving non-configural processing (Task 1B). None of the 2-way or 3-way interactions were significant. TPB and CombFB did not have a differential effect in improving r_a in the two tasks. On the whole, the results did not support H1e.

Table 5.18

Results of Judgment Achievement (r_a) for the TPF and CombFB Groups across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Feedback	0.050	1	0.050	0.483	0.489
Task	0.339	1	0.339	3.301	0.073
Feedback*Task	0.050	1	0.050	0.490	0.486
Error	8.739	85	0.103		
Within subjects:					
Block	3.596	1	3.596	73.963	0.000
Block *Feedback	0.006	1	0.006	0.117	0.733
Block*Task	0.003	1	0.003	0.055	0.815
Block*Feedback*Task	0.008	1	0.008	0.156	0.694
Error	4.132	85	0.049		

Panel B: Average Judgment Achievement (r_a) of the TPF and CombFB Groups for Two Tasks in Two Blocks

Task 1A

	TPF (n = 21)	**CombFB (n = 22)**
Block 1	0.5520	0.5496
Block 2	0.7295	0.7512
Difference	0.1775	0.2016

Task 1B

	TPF (n = 24)	**CombFB (n = 22)**
Block 1	0.6525	0.6085
Block 2	0.7782	0.7382
Difference	0.1257	0.1297

5.4 Test of Hypotheses - Self-insight

Feedback may not improve judgment accuracy in the same manner and magnitude for all individuals. This is particularly the case when involving CFB. CFB would be effective in improving judgment accuracy for individuals who know little about their own judgment/decision rules, or the weights they assigned to each factor (or cue) in their judgments or decisions. That is, individuals who have poor self-insight in their judgment processes, models, or weightings. The following hypotheses attempt to analyse the effect of self-insight on the effectiveness of TPF, CFB and CombFB.

After completing the first block of trials of 16 cases in the risk assessment task, subjects were asked to assign a weight (from 0 to 100) to each of the five procedures (cues) provided to them in each of the 16 cases. For each subject, the subjective cue weights were converted to a total of 100% by dividing the weight assigned to each cue by the sum of the weights of the five cues. For each subject, the actual judgment weights were estimated based on the response to the first block of 16 cases. As only the subjective weights on the five main effects were solicited for each subject, the statistical cue weights of the five cues (i.e., only the five main effects) were converted to a total of 100% by the same procedures as for the subjective weights mentioned above. Then a correlation was calculated for each subject between his/her subjective weights and statistically modeled weights. As Task 1A and Task 1B are slightly different in their cue weights and of different complexity, I used the respective median of the correlation of all subjects in each task as the cut-off point. A subject is classified as having high (low) self-insight if his/her correlation is equal to or above (below) the median.

5.4.1 Hypothesis H2a

H2a predicts that CFB is more effective in improving judgment performance for individuals with low self-insight than for those with high self-insight in both tasks of low and high complexity. To test this hypothesis, a repeated measures ANOVA model was run. The dependent variable was the judgment accuracy (MAE) of subjects in the CFB groups in both tasks in the two blocks of trials. The independent variables were self-insight, task complexity (both between-subjects factors) and block as a within-subjects factor. The results of the ANOVA and mean judgment accuracy are presented in Table 5.19.

Table 5.19

Results of Judgment Accuracy (MAE) for CFB and Self-insight in Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Self-insight	86.94	1	86.94	2.808	0.101
Task	18.09	1	18.09	0.584	0.449
Self-insight*Task	4.03	1	4.03	0.130	0.720
Error	1,300.52	42	30.97		
Within subjects:					
Block	159.96	1	159.96	28.873	0.000
Block *Self-insight	31.02	1	31.02	5.599	0.023
Block*Task	56.73	1	56.73	10.239	0.003
Block*Self-insight*Task	11.89	1	11.89	2.146	0.150
Error	232.69	42	5.54		

Panel B: Average Judgment Accuracy (MAE) of the CFB Condition across Two Tasks

Task 1A

	Subjects with Low Self-insight (n = 12)	Subjects with High Self-insight (n = 11)
Block 1	20.00	16.59
Block 2	13.91	14.26
Difference	6.09	2.33

Task 1B

	Subjects with Low Self-insight (n = 12)	Subjects with High Self-insight (n = 11)
Block 1	17.24	14.43
Block 2	15.73	13.01
Difference	1.51	1.42

There were two significant 2-way interaction effects. Both Block * Self-insight ($F_{(1,42)} = 5.599$, $p = 0.023$) and Block * Task ($F_{(1,42)} = 10.239$, $p = 0.003$) interactions were significant. These two significant 2-way interactions seemed to be driven by the significant improvement in judgment accuracy for subjects with low self-insight in Task 1A (see Table 5.19). Pair-wise comparisons using a Tukey HSD test indicated that the group with low self-insight, who completed Task 1A, exhibited a significantly higher improvement in accuracy than did the other three groups ($p < 0.05$). Meanwhile, the other three groups exhibited no significant difference in judgment accuracy improvement. The results suggested a possible 3-way interaction. However, the ANOVA results did not show a significant 3-way interaction ($F_{(1,42)} = 2.146$, $p = 0.150$). This was possibly due to the extremely significant Block main effect ($p = 0.000$). ANOVA models normally attribute variances to main effects before attributing variances to interaction effects. This is particularly the case for higher-order interactions. Another possible reason for failing to have a significant Block * Self-insight * Task 3-way interaction may be the relatively small sample size. There were only 11 to 12 subjects in each cell upon classifying subjects into high or low self-insight groups (see Panel B of Table 5.19).

The significant Block * Task 2-way interaction can also be explained by the above pair-wise comparisons. It was mainly driven by the fact that low self-insight subjects who completed Task 1A exhibited significantly higher improvement than did the subjects who completed Task 1B. The mean improvement in the judgment accuracy for low and high self-insight subjects who completed Task 1A was 6.09

and 2.33, respectively. For Task 1B, the mean improvement in judgment accuracy for subjects with low and high self-insight was 1.51 and 1.42, respectively. The difference was quite marginal. The significant Block * Task interaction and Tukey HSD test result for Task 1A suggested that CFB was significantly more effective in improving judgment accuracy for low self-insight subjects than for their high self-insight counterparts. However, CFB led to no significant difference in the improvement of judgment accuracy between low and high self-insight subjects in Task 1B (a task of low complexity).

The repeated measures ANOVA results with judgment achievement (r_a) as the dependent variable (presented in Table 5.20) were very similar to those discussed above, except that the Block * Self-insight * Task 3-way interaction was marginally significant ($p < 0.085$). The comparison examining improvement in judgment achievement (r_a) across the four groups of subjects (high or low self-insight; completion of Task 1A or Task 1B) had similar results as those of judgment accuracy. The effect of CFB in improving judgment achievement (r_a) for low and high self-insight subjects in Task 1A was 0.3116 and 0.0606, respectively. Low self-insight subjects in the CFB group who completed Task 1A showed higher improvement in judgment achievement (r_a) than did their high self-insight counterparts. For Task 1B, the difference in improvement in judgment achievement (r_a) for low and high self-insight subjects was small (see Table 5.20).

The above results only partially support H2a. The results for subjects who completed a task of high complexity (Task 1A) support H2a. Subjects who completed Task 1A with low self-insight and received CFB, exhibited greater improvement in judgment accuracy (MAE) and judgment achievement (r_a) than did their high self-insight counterparts. That is, self-insight moderated the effectiveness of CFB in improving judgment accuracy and judgment achievement in a task of high complexity involving configural cue processing. There was no such significant moderating effect in improving judgment accuracy or judgment achievement in a task of low complexity.

Table 5.20

Results of Judgment Achievement (r_a) for CFB and Self-insight in Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Self-insight	0.254	1	0.254	2.453	0.387
Task	0.079	1	0.079	0.764	0.125
Self-insight*Task	0.061	1	0.061	0.585	0.449
Error	4.349	42	0.104		
Within subjects:					
Block	0.665	1	0.665	23.756	0.000
Block *Self-insight	0.419	1	0.419	14.982	0.000
Block*Task	0.221	1	0.221	7.907	0.007
Block*Self-insight*Task	0.087	1	0.087	3.105	0.085
Error	1.175	42	0.028		

Panel B: Average Judgment Achievement (r_a) of the CFB group across Two Tasks

Task 1A

	Subjects with Low Self-insight (n = 12)	**Subjects with High Self-insight (n = 11)**
Block 1	0.4243	0.6015
Block 2	0.7359	0.6621
Difference	0.3116	0.0606

Task 1B

	Subjects with Low Self-insight (n = 12)	**Subjects with High Self-insight (n = 11)**
Block 1	0.5623	0.7040
Block 2	0.6429	0.7068
Difference	0.0806	0.0028

5.4.2 Hypothesis H2b

H2b predicts that CombFB is more effective in improving judgment performance than TPF alone for subjects with low self-insight in both tasks of low and high complexity. A repeated measures ANOVA was run to test the hypothesis with low self-insight subjects in the TPF and CombFB groups in both tasks. The dependent variable was judgment accuracy (MAE) and the independent variables were task complexity and feedback factors (manipulated between-subjects) and block was a within-subjects factor. To support the hypothesis, there would be a significant Block * Feedback 2-way interaction. The results of the ANOVA and average judgment accuracy are presented in Table 5.21.

The only significant effect was the Block main effect, indicating that subjects as a whole had a significant improvement in judgment accuracy (MAE) in the second block of trials from the first block after having received the relevant feedback. However, Task or Feedback did not show any significant main effect. The average improvement in judgment accuracy for the two tasks under the two feedback conditions is shown in Panel B of Table 5.21, which reveals that the average improvement in judgment accuracy was in the expected direction of the hypothesis, but did not achieve the required significant level. CombFB (TPF) led to an average improvement in judgment accuracy in Task 1A and Task 1B of 7.22 and 5.06 (4.50 and 3.80), respectively. One possible reason might be the relatively small sample size. There were only 10 to 13 subjects with low self-insight in each task and feedback condition.

The ANOVA and average judgment improvement for judgment achievement (r_a) are presented in Table 5.22. The results were very similar to those presented above. Again, the average improvement in judgment achievement between the two feedback conditions was in the expected direction, but failed to achieve the required significant level. On the whole, the results did not support H2b.

Table 5.21

Results of Judgment Accuracy (MAE) for TFB and CombFB with Low Self-insight Subjects across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	3.87	1	3.87	0.117	0.734
Feedback	0.01	1	0.01	0.000	0.990
Task*Feedback	13.90	1	13.90	0.420	0.521
Error	1,324.64	40	33.12		
Within subjects:					
Block	575.52	1	575.52	48.224	0.000
Block *Task	11.08	1	11.08	0.929	0.341
Block*Feedback	21.54	1	21.54	1.805	0.187
Block*Task*Feedback	2.86	1	2.86	0.240	0.627
Error	477.37	40	11.93		

Panel B: Average Judgment Accuracy (MAE) of the Feedback Conditions across Two Tasks for Low Self-insight Subjects

Task 1A

	TPF (n = 10)	**CombFB (n = 11)**
Block 1	18.06	18.64
Block 2	13.56	11.42
Difference	4.50	7.22

Task 1B

	TPF (n = 13)	**CombFB (n = 10)**
Block 1	16.49	17.94
Block 2	12.69	12.88
Difference	3.80	5.06

Table 5.22

Results of Judgment Achievement (r_a) for TFB and CombFB with Low Self-insight Subjects across Two Tasks

Panel A: ANOVA Results

	SS	df	MS	F ratio	P
Between subjects:					
Task	0.105	1	0.105	0.962	0.332
Feedback	0.001	1	0.001	0.010	0.922
Task*Feedback	0.044	1	0.044	0.406	0.528
Error	4.353	40	0.109		
Within subjects:					
Block	3.087	1	3.087	67.157	0.000
Block *Task	0.103	1	0.103	2.235	0.143
Block*Feedback	0.056	1	0.056	1.212	0.278
Block*Task*Feedback	0.000	1	0.000	0.005	0.945
Error	1.839	40	0.046		

Panel B: Average Judgment Achievement (r_a) of the Feedback Conditions across Two Tasks for Low Self-insight Subjects

Task 1A

	TPF (n = 10)	CombFB (n = 11)
Block 1	0.4498	0.4817
Block 2	0.7156	0.7820
Difference	0.2658	0.3003

Task 1B

	TPF (n = 13)	CombFB (n = 10)
Block 1	0.6063	0.5624
Block 2	0.7497	0.7286
Difference	0.1434	0.1662

5.5 Testing of Hypotheses – Transfer of Learning

The previous analyses focused on the investigation of the improvement in judgment performance within the same task under the five feedback conditions across two blocks. According to popular learning theories, learning is more effective if it can be applied to a similar or new task. That is, the learning in one task can be transferred to another. This aspect of learning is important as the business environment is always changing and auditors are constantly required to conduct audits in new business environments and transactions. It would be impossible to provide training for all these new business practices and transactions to auditors, especially on an *ex ante* basis. Therefore, the transfer of learning benefits of feedback are of interest to auditors.

5.5.1 Hypothesis H3a

H3a hypothesises that subjects practised in a configural cue processing task (Task 1A) would perform better in a subsequent task (Task 2), which requires similar information processing than those who practised in a non-configural task (Task 1B). Judgment performance on Task 2 was compared between subjects who completed Task 1A and Task 1B in the previous two blocks of trials. An independent samples t-test was conducted to compare judgment accuracy (MAE) in Task 2 for two groups (practised in a configural cue processing task or not). There was a significant difference in judgment accuracy in Task 2 between the two groups of subjects (t_{222} =

3.560, $p = 0.000$). A t-test comparing judgment achievement (r_a) was also significant ($t_{222} = 2.945$, $p = 0.004$). The results supported H3a.

The above results only indicate that subjects who completed a configural cue processing task had better judgment accuracy and judgment achievement in Task 2 than those who completed the non-configural case. However, the results did not pinpoint which feedback condition contributes to the difference. Therefore, two One-way ANOVAs were conducted, one with judgment accuracy and the other with judgment achievement in Task 2 as the dependent variables and the 10 [2 (levels of task complexity) X 5 (feedback treatments)] groups as independent variables. The One-way ANOVA results are presented in Table 5.23.

The ANOVA models for judgment accuracy (MAE) and judgment achievement (r_a) were significant ($F_{(9, 214)} = 3.333$, $p = 0.001$) and ($F_{(9, 214)} = 2.953$, $p = 0.003$), respectively. The results indicated that the ten groups of subjects had significantly different judgment accuracy and judgment achievement in Task 2.

Table 5.23

One-way ANOVA Results Comparing Judgment Performance in Task 2 across Ten Feedback Groups

Panel A: One-way ANOVA Results on Judgment Accuracy (MAE) in Task 2 with 10 Feedback Groups

Source	SS	df	MS	F ratio	P
Between Groups	473.39	9	52.60	3.333	0.001
Within Groups	3,376.88	214	15.78		
Total	3,850.27	223			

Panel B: One-way ANOVA Results on Judgment Achievement (r_a) in Task 2 with 10 Feedback Groups

Source	SS	df	MS	F ratio	P
Between Groups	0.834	9	0.093	2.953	0.003
Within Groups	6.714	214	0.032		
Total	7.548	223			

The between-groups differences were significant indicating that there were differences in judgment accuracy in Task 2 among the ten groups with various feedback and levels of task complexity conditions. Results of pair-wise multiple comparisons (using the Tukey HSD test) of the judgment accuracy (MAE) in Task 2 indicated that the between-groups differences were mainly due to significantly better judgment accuracy of the groups who completed the configural cue processing task (high in task complexity) and received CombFB (see Table 5.24 for the mean judgment accuracy for each group and their pair-wise comparisons). Thus, practice in a task of high complexity and receiving CombFB, facilitated the transfer of knowledge gained in the process to another task requiring similar cue processing knowledge. Subjects who completed the configural cue processing task and received CombFB also had significantly higher judgment accuracy in Task 2 than those who completed the same configural cue processing task but received NFB ($p < 0.05$) and OFB ($p < 0.1$). For subjects who completed Task 1A there was no significant difference between those who received TPF, CFB, and CombFB.

Results of the pair-wise comparisons (using the Tukey HSD test) of the judgment achievement (r_a) among the ten groups were very similar to those discussed above and are presented in Table 5.25. Subjects in the CombFB group who completed a configural cue processing task performed significantly better in terms of judgment achievement in Task 2 than subjects who completed a non-configural cue processing task. For the five groups of subjects who completed a configural cue processing task, only the CombFB group had significantly higher judgment achievement than the NFB group ($p < 0.05$) and was marginally higher than the OFB group ($p < 0.1$). There was no significant difference among TPF, CFB and CombFB groups.

On the whole, the results supported H3a. This was particularly the case if subjects practised a task involving configural cue processing and received CombFB.

Table 5.24
Mean Judgment Accuracy (MAE) in Task 2 for Five Feedback Treatments across Two Tasks

Panel A: Mean Judgment Accuracy

	Task 1A					Task 1B				
Group	NFB 1	OFB 2	TPF 3	CFB 4	CombFB 5	NFB 6	OFB 7	TPF 8	CFB 9	CombFb 10
Mean	17.59	17.20	14.64	15.84	13.69	18.49	17.33	17.60	17.26	18.07

Panel B: Pair-wise Comparisons with Tukey HSD Test of the Difference in Mean Judgment Accuracy (Column minus Row)

Group	1	2	3	4	5	6	7	8	9	10
1	0	-0.39	-2.95	-1.75	-3.90**	0.90	-0.26	0.01	-0.33	0.48
2		0	-2.56	-1.36	-3.51*	1.29	0.13	0.40	0.06	0.87
3			0	1.20	-0.95	3.85**	2.69	2.96	2.62	3.43
4				0	-2.15	2.65	1.49	1.76	1.42	2.23
5					0	4.80***	3.64*	3.91**	3.57*	4.38***
6						0	-1.16	-0.89	-1.23	-0.42
7							0	0.27	-0.07	0.74
8								0	-0.34	0.47
9									0	0.81
10										0

* sig. at 0.1 level
** sig. at 0.05 level
*** sig. at 0.01 level

Table 5.25
Mean Judgment Achievement (r_a) in Task 2 for Five Feedback Treatments across Two Tasks

Panel A: Mean Judgment Achievement

	Task 1A					Task 1B				
Group	NFB 1	OFB 2	TPF 3	CFB 4	CombFB 5	NFB 6	OFB 7	TPF 8	CFB 9	CombFb 10
Mean	0.5090	0.5188	0.5701	0.5418	0.6319	0.4760	0.5203	0.5133	0.5178	0.4705

Panel B: Pair-wise Comparisons with Tukey HSD Test of the Difference in Mean Judgment Achievement (Column minus Row)

Group	1	2	3	4	5	6	7	8	9	10
1	0	0.0098	0.0611	0.0328	0.1229**	-0.0330	0.0113	0.0043	0.0088	-0.0385
2		0	0.0513	0.0230	0.1131*	-0.0428	0.0015	-0.0055	-0.0010	-0.0483
3			0	-0.0283	0.0618	-0.0941	-0.0498	-0.0568	-0.0523	-0.0996
4				0	0.0901	-0.0658	-0.0215	-0.0285	-0.0240	-0.0713
5					0	-0.1559***	-0.1116*	-0.1186**	-0.1141**	-0.1614***
6						0	0.0443	0.0373	0.0418	0.0055
7							0	-0.0070	-0.0025	-0.0498
8								0	0.0045	-0.0428
9									0	-0.0473
10										0

* sig. at 0.1 level
** sig. at 0.05 level
*** sig. at 0.01 level

5.5.2 Hypothesis H3b

H3b predicts that subjects who completed a task involving configural cue processing (Task 1A) and received TPF have higher judgment performance in Task 2 than subjects who completed the same task but received NFB. As noted in Table 5.24, the mean judgment accuracy for subjects who completed Task 1A and received TPF and NFB was14.64 and 17.59, respectively. The performance for these two groups of subjects in Task 2 was in the hypothesised direction, but did not achieve the 5% significance level.

A similar result was found when comparing the judgment achievement (r_a) of these two groups of subjects (see Table 5.25). On the whole, the results did not support H3b.

5.5.3 Hypothesis H3c

H3c predicts that subjects who completed a task involving configural cue processing (Task 1A) and received CFB have higher judgment performance in Task 2 than subjects who completed the same task but received NFB. As noted in Table 5.24, subjects who received CFB and completed a task involving configural cue processing did not perform significantly better than their NFB counterparts in terms of judgment accuracy (MAE) in Task 2. The average judgment accuracy for NFB and CFB subjects was 17.59 and 15.84, respectively. But the differences were not statistically significant.

The result comparing subjects' judgment achievement (r_a) were similar to the above. The average judgment achievement for NFB and CFB subjects was 0.5090 and 0.5418, respectively (see Table 5.25). On the whole, the results did not support H3c.

5.5.4 Hypothesis H3d

H3d predicts that subjects who completed a task involving configural cue processing (Task 1A) and received CombFB have higher judgment performance in Task 2 than subjects who completed the same task but received NFB. The average judgment accuracy in Task 2 for NFB and CombFB subjects was 17.59 and 13.69, respectively (see Table 5.24). The difference was significant ($p < 0.05$). Subjects who received CombFB and completed a task involving configural cue processing performed better than their NFB counterparts in terms of judgment accuracy in Task 2.

The difference in judgment achievement (r_a) between the two feedback groups (see Table 5.25) was also significant ($p < 0.05$). The result was in line with that of judgment accuracy. The results supported H3d.

5.6 Additional Analysis

Task 1A requires configural cue processing in making the appropriate judgment. Specifically, two of the audit procedures (cues) were substitutable for each other. They interact in a compensatory ordinal form in an ANOVA model (see the Research Method Section for details). However, whether feedback would promote configural cue processing has not been studied. One of the objectives of this research is to address this issue. This study investigated the effect of five feedback conditions in promoting configural cue processing in an audit risk assessment task. The five feedback conditions were NFB, OFB, TPF, CFB, and CombFB. Task 1B does not involve configural cue processing and, therefore, would be excluded from this part of the analysis.

Recall that each subject completed two blocks of trials of 16 cases, each of which represented a half-replication of a 2^5 factorial design. Between the two blocks of trials, subjects in the TPF, CFB, and CombFB groups received the appropriate feedback before proceeding to the second block of 16 cases, which was the other half of the 2^5 factorial design not used in the first block. Subjects in the OFB group received outcome feedback after they completed each of the 16 cases in the first block of trials. Subjects in the NFB group received no feedback information in the two blocks of trials.

For each subject and for each block of trials, I modeled the judgment weights on each of the five cues and each of the 10 two-way interactions. The weight attributed to the interaction term of the two substitutable cues was the measure of the extent of the configural cue processing for that subject in that block of trials. For each subject, the extent of configural cue processing in Task 1A was measured twice, once for each block of trials. The difference in the extent of the configural cue processing between the two blocks of trials was also computed and analysed.

5.6.1 Effect of Different Feedback on Promoting Configural Cue Processing

The average extent of configurality in each block and their difference are presented in Table 5.26. The ANOVA model comparing the extent of configural cue processing in the first block of trials was not significant ($F_{(4,106)} = 0.057$, $p = 0.994$). The results indicated that there was no significant difference in the extent of configural cue processing among subjects in the five feedback groups in the first block of trials.

The One-way ANOVA on the extent of configural cue processing across the five feedback treatment groups in Block 2 was significant ($F_{(4, 106)} = 3.365$, $p = 0.012$). Pair-wise comparisons with Tukey's HSD test revealed that there were significant differences between the CombFB group with NFB or OFB groups. Subjects in the CombFB group exhibited significant increases in configural cue processing compared to those in the NFB or OFB groups who completed Task 1A. There was no significant difference in the extent of configural cue processing between the two blocks of trials among the TPF, CFB, and CombFB groups.

The number of subjects who exhibited above-criterion (i.e., 4%) weight on the interaction term in the NFB, OFB, TPF, CFB, and CombFB groups was 5, 6, 5, 4 and 5 in Block 1, and 5, 7, 15, 10, and 16 in Block 2, respectively (see Panel A of Table 5.26). As a subject can exhibit either above-criterion weighting on the interaction term or not, it was a nominal data set. A Kruskal-Wallis test was conducted to test whether subjects exhibited above-criterion weighting on the interaction term was different among the five feedback groups. The result was not significant in the first block of trials ($\chi^2 = 0.531$, $p = 0.970$), but significant in the second block of trials ($\chi^2 = 18.506$, $p = 0.001$). The Kruskal-Wallis test does not identify which groups had more subjects who exhibited above-criterion weighting. However, it appeared that the number of subjects exhibiting above-criterion weighting on the interaction term was higher in the TPF, CFB, and CombFB groups than the NFB and OFB groups.

The number of subjects who exhibited above-criterion weighting on the interaction term between the two blocks of trials was also compared with chi-square tests (see Table 5.26). Of the 111 subjects who completed Task 1A, 25 (or 22.5%) exhibited above-criterion weighting of the interaction term in the first block of trials. In the second block of trials, the number of subjects who exhibited above-criterion weighting increased to 53 (47.7%). Using the chi-square test the proportion of subjects who exhibited above-criterion weighting across the two blocks of trials was significantly higher in Block 2 ($\chi^2 = 40.577$, $p = 0.000$) than in block 1.

For each feedback treatment, the difference in the number of subjects who exhibited above-criterion weight between the two blocks of trials was also tested. It was found that there was no significant difference for the NFB ($\chi^2 = 0.001$, $p = 0.980$), and OFB ($\chi^2 = 0.830$, $p = 0.360$). On the other hand, there were significantly more subjects who exhibited above-criterion weighting on the interaction term in their judgment models in TPF ($\chi^2 = 28.831$, $p = 0.000$), CFB ($\chi^2 = 5.805$, $p = 0.016$), and CombFB ($\chi^2 = 31.829$, $p = 0.000$) groups.

The average weight attributed to the interaction term in the two blocks of trials for the five feedback conditions was also compared. The results were similar to those found in comparing the number of subjects who exhibited above-criterion weighting of the interaction term. It was found that significantly more weight was attributed to the interaction terms in the second block than the first block of trials for the TPF ($t_{20} = 2.829$, $p = 0.010$), CFB ($t_{22} = 2.201$, $p = 0.039$), and CombFB ($t_{21} = 3.828$, $p = 0.001$) groups. There was no significant difference found in the NFB ($t_{21} = 0.046$, $p = 0.964$), and OFB ($t_{22} = 0.079$, $p = 0.938$) groups in the weight attributed to the interaction term between the two blocks of trials.

The increase in weight attributed to the interaction term (configural cue combination) across the two blocks among the five feedback conditions was also compared. A One-way ANOVA was conducted with the feedback groups as the independent variable and weight attributed to the interaction term as the dependent variable. The ANOVA model was significant ($F_{(4, 106)} = 5.360$, $p = 0.001$) indicating that there was a difference in the increase in weight attributed to the interaction term between the subjects of the different feedback groups. Pair-wise comparisons with the Tukey HSD test identified that there was a significant difference in the increase in weight attributed to the interaction term between the CombFB groups and the NFB and OFB groups. These results are shown in Panel B of Table 5.26. There was no significant difference between the CombFB and TPF groups, and the CFB groups.

Table 5.26

The Extent of Configural Cue Processing in Two Blocks of Trials for Five Feedback Treatments

Panel A: Average Weight Attributed to the Configural Cue Combination

		Block 1		Block 2		Difference	
Feedback	**n**	**No. of subjects > 4%**	**All subjects mean**	**No. of subjects > 4%**	**All subjects mean**	**No. of subjects > 4%**	**All subjects mean**
NFB	22	5	3.840	5	3.777	0	-0.063
OFB	23	6	3.991	7	4.088	1	0.097
TPF	21	5	3.662	15	8.640	10***	4.978***
CFB	23	4	3.237	10	7.856	6**	4.619**
CombFB	22	5	3.428	16	11.654	11***	8.226***
Total No.	111	25		53		28	
Average			3.631		7.168		3.537

Panel B: Pair-wise Comparisons (Tukey HSD test) on Difference in Increased Weight Attributed to the Configural Cue between Two Blocks

	Difference in Mean (column minus row)				
	NFB	**OFB**	**TPF**	**CFB**	**CombFB**
NFB	0	0.16	5.04	4.682	8.289**
OFB		0	4.881	4.522	8.129**
TPF			0	-0.359	3.248
CFB				0	3.067
CombFB					0

** significant at 0.05 level

*** significant at 0.01 level

5.6.2 Further Analysis on Lens Model Parameters

Some of the earlier data analyses are focused on judgment achievement (r_a) – one of the lens model parameters. The effects of various feedback conditions on other lens model parameters such as judgment consistency (Rs), linear knowledge (G) and non-linear knowledge (C) in the task are also analysed in this section. The descriptive statistics of the lens model parameters are presented in Table 5.27.

5.6.2.1 Judgment Consistency (Rs)

Judgment consistency (Rs) measures how consistent subjects were in applying their judgment models and weights in their actual judgments. It would affect the judgment achievement and judgment accuracy. Table 5.28 presents the average Rs in two blocks of trials in each of the feedback conditions for Task 1A. TPF, CFB, and CombFB led to significant improvements in Rs between the two blocks of trials. Pair-wise comparisons (using the Tukey HSD test) of improvement in Rs between feedback conditions revealed that the TPF and CombFB groups were significantly better than the NFB and OFB groups. CFB group also showed marginal improvement in Rs over the NFB and OFB groups. There was no significant difference between the TPF, CFB, and CombFB groups. Subjects generally applied their judgment models fairly consistently in both blocks of trials. They have the required task knowledge and auditing experience to complete this familiar audit task and therefore would have high consistency.

Table 5.27

Descriptive Statistics of Lens Model Parameters

Panel A: Group means of Task 1A (Standard deviation shown in paretheses)

	Block 1						Block 2					
	NFB	OFB	TPF	CFB	CombFB	Overall	NFB	OFB	TPF	CFB	CombFB	Overall
Rs	0.81 (0.08)	0.81 (0.06)	0.83 (0.03)	0.82 (0.05)	0.83 (0.04)	0.82 (0.05)	0.83 (0.03)	0.81 (0.04)	0.87 (0.03)	0.86 (0.04)	0.88 (0.03)	0.85 (0.04)
G	0.67 (0.17)	0.64 (0.17)	0.68 (0.18)	0.66 (0.13)	0.73 (0.06)	0.68 (0.15)	0.61 (0.33)	0.66 (0.17)	0.78 (0.04)	0.75 (0.06)	0.79 (0.04)	0.73 (0.18)
C	0.16 (0.35)	0.20 (0.34)	0.22 (0.37)	0.14 (0.38)	0.16 (0.34)	0.18 (0.35)	0.17 (0.40)	0.23 (0.36)	0.45 (0.24)	0.40 (0.29)	0.55 (0.29)	0.36 (0.36)

Panel B: Group Mean in Task 1B

	Block 1						Block 2					
	NFB	OFB	TPF	CFB	CombFB	Overall	NFB	OFB	TPF	CFB	CombFB	Overall
Rs	0.86 (0.05)	0.85 (0.06)	0.85 (0.05)	0.84 (0.11)	0.86 (0.07)	0.85 (0.07)	0.87 (0.04)	0.91 (0.06)	0.91 (0.04)	0.87 (0.05)	0.91 (0.04)	0.89 (0.05)
G	0.69 (0.21)	0.72 (0.14)	0.72 (0.14)	0.69 (0.18)	0.68 (0.17)	0.70 (0.17)	0.63 (0.24)	0.78 (0.10)	0.80 (0.10)	0.73 (0.16)	0.79 (0.08)	0.75 (0.16)

Table 5.28
Judgment Consistency (Rs) for Each Feedback Treatment for Task 1A

Panel A: Average Judgment Consistency (Rs) in Both Blocks of Trials in Task 1A (with configural cues)

Feedback	n	Mean Rs # in 1st Block	Mean Rs in 2nd Block	Difference in Mean Rs
NFB	22	0.8147	0.8270	0.0123
OFB	23	0.8059	0.8145	0.0086
TPF	21	0.8256	0.8718	0.0462***
CFB	23	0.8189	0.8605	0.0416***
CombFB	22	0.8288	0.8764	0.0476***
Average		0.8186	0.8496	0.0310

no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

Panel B: Pairwise Comparison Using the Tukey HSD Test of the Mean Difference in Rs between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	-0.0037	0.0339**	0.0293*	0.0353**
OFB		0	0.0376**	0.0330*	0.0390**
TPF			0	-0.0046	0.0014
CFB				0	0.0060
CombFB					0

* sig. at 0.1 level
** sig. at 0.05 level
*** sig., at 0.01 level

Table 5.29
Judgment Consistency (Rs) for Each Feedback Treatment for Task 1B

Panel A: Average Judgment Consistency (Rs) in Both Blocks of Trials in Task 1B (with non-configural cues)

Feedback	n	Mean Rs # in 1st Block	Mean Rs in 2nd Block	Difference in Mean Rs
NFB	22	0.8577	0.8662	0.0085
OFB	22	0.8517	0.9053	0.0536***
TPF	24	0.8505	0.9115	0.0610***
CFB	23	0.8389	0.8732	0.0344*
CombFB	22	0.8557	0.9141	0.0584***
Average		0.8508	0.8942	0.0434

no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

Panel B: Pairwise Comparison Using the Tukey HSD Test of the Mean Difference in Rs between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	0.0451**	0.0525**	0.0259	0.0499**
OFB		0	0.0074	-0.0192	0.0048
TPF			0	-0.0266	-0.0026
CFB				0	0.0240
CombFB					0

* sig. at 0.1 level
** sig. at 0.05 level
*** sig. at 0.01 level

For Task 1B, the results indicated that OFB, TPF and CombFB led to significant improvements in Rs between the two blocks of trials and CFB had only marginal improvements (see Table 5.29). The comparison of, the improvement across the five feedback conditions shows OFB, TPF, and CombFB had significantly higher improvements in Rs compared with the NFB group.

5.6.2.2 Knowledge of the Linear Component of the Task (G)

Subjects' knowledge of the linear component of the task (G) will also affect their judgment performance. This knowledge would enable subjects to identify the importance and validity of each cue. The results of G for Task 1A and Task 1B are presented in Tables 5.30 and 5.31. TPF, CFB, and CombFB led to significant increases in G in the second block of trials within their own groups. Also, these three feedback conditions led to significant increases in G when compared with the NFB and OFB groups.

The results for Task 1B were similar to that of Task 1A with respect to TPF and CombFB. Both feedback conditions led to significant increases in G between blocks and within their own groups. In contrast to Task 1A, OFB was significant in improving G between the two blocks of trials in Task 1B. On the other hand, CFB had only a marginal effect in improving G between the blocks of trials. OFB, TPF and CombFB led to significant improvement in G compared to the NFB condition.

Table 5.30
Knowledge of the Linear Component of Task (G) for Each Feedback Treatment for Task 1A

Panel A: Average G in Both Blocks of Trials in Task 1A (with configural cues)

Feedback	n	Mean G # in 1st Block	Mean G in 2nd Block	Difference in Mean G
NFB	22	0.6737	0.6070	-0.0667
OFB	23	0.6396	0.6626	0.0230
TPF	21	0.6824	0.7790	0.0966***
CFB	23	0.6632	0.7522	0.0890***
CombFB	22	0.7278	0.7977	0.0699***
Average		0.6768	0.7189	0.0421

no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

Panel B: Pairwise Comparison Using the Tukey HSD Test of the Mean Difference in G between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	0.0897	0.1633**	0.1557**	0.1366**
OFB		0	0.0736	0.0669	0.0469
TPF			0	-0.0076	-0.0267
CFB				0	-0.0191
CombFB					0

** sig. at 0.05 level
*** sig. at 0.01 level

Table 5.31
Knowledge of the Linear Component of Task (G) for Each Feedback Treatment for Task 1B

Panel A: Average G in Both Blocks of Trials in Task 1B (with non-configural cues)

Feedback	n	Mean G # in 1st Block	Mean G in 2nd Block	Difference in Mean G
NFB	22	0.6853	0.6319	-0.0534
OFB	22	0.7164	0.7784	0.0620**
TPF	24	0.7167	0.8022	0.0855***
CFB	23	0.6888	0.7324	0.0436*
CombFB	22	0.6829	0.7904	0.1075***
Average		0.6983	0.7479	0.0496

no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

Panel B: Pairwise Comparison Using the Tukey HSD Test of the Mean Difference in G between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	0.1154**	0.1389***	0.0970	0.1609***
OFB		0	0.0235	-0.0184	0.0455
TPF			0	-0.0419	0.0220
CFB				0	0.0639
CombFB					0

* sig. at 0.1 level
** sig. at 0.05 level
*** sig. at 0.01 level

5.6.2.3 Knowledge of the Non-linear Component of the Task (C)

Only Task 1A involved configural cue processing. Therefore, the effect of the knowledge of the non-linear component of the task (C) was also computed and compared across the treatment groups. The results are presented in Table 5.32. As shown in Table 5.32, TPF, CFB, and CombFB led to significant improvement in C in the second block of trials. The improvement was particularly noticeable for the CombFB condition ($p < 0.01$). NFB and OFB did not lead to a significant improvement in C between the two blocks of trials.

However, when comparing the effectiveness in improving C between the two blocks of trials among different feedback groups, only the CombFB group exhibited marginal improvement over the NFB and OFB groups even though the difference between the blocks was quite large (see Panel A of Table 5.32). The lack of significant difference of improvement of C between the blocks for various feedback conditions might be due to the large variations in distribution or standard deviations in C. The standard deviations ranged from 0.3627 for OFB group to 0.3648 for the CFB group.

Table 5.32
Knowledge of the Non-linear Component of Task (C) for Each Feedback Treatment for Task 1A

Panel A: Average C in Both Blocks of Trials in Task 1A (with configural cues)

Feedback	n	Mean G # in 1st Block	Mean G in 2nd Block	Difference in Mean G
NFB	22	0.1609	0.1681	0.0072
OFB	23	0.1973	0.2252	0.0279
TPF	21	0.2160	0.4524	0.2364**
CFB	23	0.1443	0.4033	0.2590**
CombFB	22	0.1627	0.5479	0.3852***
Average		0.1758	0.3577	0.1819

no significant difference among the five groups in the 1st block (i.e., before treatment / feedback)

Panel B: Pairwise Comparison Using the Tukey HSD test of the Mean Difference in C between the First and Second Blocks of Trials

	Difference in Mean (Column minus Row)				
	NFB	OFB	TPF	CFB	CombFB
NFB	0	0.0207	0.2292	0.2518	0.3780*
OFB		0	0.2085	0.2311	0.3573*
TPF			0	0.00226	0.1488
CFB				0	0.1262
CombFB					0

* sig. at 0.1 level
** sig. at 0.05 level
*** sig. at 0.01 level

5.6.3 Behavioral Responses to Different Types of Feedback

Differences in the types of feedback provided to subjects also affected their responses to the three self-report measures used to assess their perceptions and reactions. Specifically, subjects were asked to answer 12 questions on their perceptions of the understandability, current helpfulness, and future helpfulness of the feedback they received. The three scales (the 12 questions) were developed and used by Balzer et al. (1992). As the contents of feedback for Task 1A and Task 1B were quite different, the responses were analysed separately for each task. One-way ANOVAs were conducted to compare the responses to the three scales in each type of feedback. The NFB groups did not receive any feedback and therefore were not included in this part of analysis.

For Task 1A, the ANOVA for mean responses to the understandability of feedback was significant ($F_{(3,85)} = 6.481$, $p = 0.001$). Pair-wise comparisons with the Tukey HSD test revealed that subjects in TPF and CombFB groups had significantly (marginally) higher ratings on the understandability than the OFB (CFB) group. The ANOVA model for current helpfulness was also significant ($F_{(3,85)} = 8.002$, $p = 0.000$) and the pair-wise comparisons results were very similar to those of understandability. Future helpfulness of the feedback was also significant ($F_{(3,85)} = 6.569$, $p = 0.000$). Pair-wise comparisons indicated that the TPF and CombFB groups had significantly higher ($p < 0.01$) rating than subjects in the OFB groups did. However, there was no significant difference between the TPF, CFB, and CombFB groups.

For Task 1B, the ANOVA for understandability was significant ($F_{(3, 87)} = 5.629$, $p = 0.001$). Pair-wise comparisons with the Tukey HSD test revealed that the subjects in the TPF and CombFB groups had significantly higher ($p < 0.05$) ratings than the OFB and CFB groups. For the current helpfulness, the ANOVA model comparing the ratings between the four feedback groups was significant ($F_{(3, 87)} = 5.875$, $p = 0.001$). Pair-wise comparisons with the Tukey HSD test had similar results to the understandability of feedback. The ANOVA for the future helpfulness of feedback was also significant ($F_{(3, 87)} = 5.424$, $p = 0.001$). Pair-wise comparisons with the Tukey HSD test also revealed the same pattern of results as the understandability and the current helpfulness. TPF and CombFB groups had significantly higher ($p < 0.05$) ratings than the OFB and CFB groups.

On the whole, the results indicated that subjects who received either TPF or CombFB had significantly higher ratings on the three scales (understandability, current helpfulness, and future helpfulness) than subjects who received OFB and CFB. This was the case even if CFB (OFB) led to significantly better judgment performance in Task 1A (Task 1B). OFB and CFB might be more difficult to understand and therefore the subjects' perception of them would not be as favourable as TPF or CombFB. It seems that TPF component of CombFB was the important element leading to a favourable perception of CombFB.

CHAPTER 6

DISCUSSION OF RESULTS

In this chapter, the results presented in Chapter 5 are discussed and compared with previous studies. The discussion is focused on (i) the effectiveness of the five feedback conditions in improving judgment performance in the two audit tasks at the two levels of task complexity; (ii) the moderating effect of self-insight on the effectiveness of TPF, CFB, and CombFB; (iii) the effect of the five feedback conditions on the transfer of learning to another task; and (iv) the results of the additional analysis of the lens model statistics.

This study found that there was a significant difference in subjects' judgment accuracy between tasks of high and low complexity. Subjects performed significantly better in a low complexity task. Also, the various feedback conditions had significantly different effects on subjects' judgment accuracy. Specifically, subjects in the OFB, TPF, CFB, and CombFB groups had significantly higher judgment accuracy than their NFB counterparts. It should be noted that the effectiveness of various feedback conditions in improving judgment accuracy was different for tasks of high and low complexity. In a task of high complexity (Task 1A), the TPF, CFB, and CombFB groups demonstrated a significantly higher improvement in judgment accuracy than the NFB and OFB groups. On the other hand, in a task of low complexity (Task 1B), the OFB, TPF, and CombFB groups showed a significant improvement in judgment accuracy compared to the NFB group.

6.1 The Effectiveness of the Five Feedback Conditions on Judgment Performance

6.1.1 Outcome Feedback

This study found that OFB was more effective in improving judgment performance (in terms of accuracy and judgment achievement) in a non-configural task (low complexity) than in a configural task (high complexity), which involved configural cue processing. The result that OFB leads to a significant improvement in judgment performance in a low complexity task is in line with more recent research involving accounting related tasks, such as financial distress prediction (Hirst et al., 1999; Tuttle and Stocks, 1998) and performance evaluation of auditors (Hirst and Luckett 1992). While overall there have been inconsistent and inconclusive results in the accounting and psychology literatures on the effectiveness of OFB in improving judgment performance, Kluger and DeNisi (1996), after an extensive review of the psychology literature and a meta-analysis on feedback intervention (OFB), concluded that outcome feedback, on average, improved judgment performance, but over one third of the results showed decreased performance. The authors suggested that task properties might moderate the effect, and that simple-task performance benefited from OFB more than complex-task performance. The results in this study provide strong support for this suggestion. The effect of OFB was moderated by the task complexity. Specifically, OFB was more effective in improving both measures of judgment performance in a task of low complexity than in a high complexity task.

The results on OFB support the arguments of Hirst and Luckett (1992), that OFB can be effective if (i) the task is simple and of high predictability; (ii) the task is familiar to the judges; and (iii) the judges possess sufficient relevant domain- and task-specific knowledge.

6.1.2 Task Properties Feedback

The study found that TPF was effective in improving judgment performance in tasks of both low and high complexity. This result is consistent with most of the previous studies in accounting, (Kessler and R. Ashton, 1981; Hirst and Luckett, 1992; Hirst et al., 1999; Tuttle and Stocks, 1998) psychology (Balzer et al., 1992; Balzer et al., 1994), and other disciplines (Remus et al., 1996). After an extensive review of the feedback literature, Balzer et al., (1989) concluded that TPF was probably the most effective and robust feedback to improve judgment performance. Most previous feedback studies involved tasks of low complexity or required subjects to process the cues in a linear manner. The results of this study extended the knowledge of the effectiveness of TPF in improving judgment performance in a task of high complexity that involved configural cues.

6.1.3 Cognitive Feedback

I found that CFB is more effective in improving judgment performance in a task that does involve configural processing than in one that does not. Specifically, I found that cognitive feedback improves performance for the configural case (Task 1A) but not for the non-configural case (Task 1B).

The results of this study on the effectiveness of CFB in a low complexity task, were consistent with those of Kessler and R. Ashton (1981) who found no significant effect of CFB in improving the accuracy of bond rating judgments. It should be noted that the bond-rating task was a task of relatively low complexity. First, it involved only three cues (accounting ratios). Second, the cue-criterion relationship was linear. Third, the optimal judgment policy of the bond-rating task assigned equal weight to the three accounting ratios (Libby 1985, p.131). The results that CFB was not effective in improving judgment performance in a non-configural task were also consistent with the findings of most psychology studies (e.g., Balzer et al. 1992; Balzer et al. 1994; Harmon and Rohrbaugh 1990). As most psychology studies used simple generic tasks "at the low ends of the dimensions of complexity" (Balzer et al. 1989, p.428), they might not be the most appropriate tasks for CFB to be effective in improving judgment performance.

In line with the results of most psychology studies and those of Kessler and R. Ashton (1981), it was found that CFB was not effective and was significantly inferior to TPF in improving judgment performance in a task of low complexity (Task 1B). However, TPF and CFB were both effective in improving judgment performance in a task of high complexity involving configural cue processing. Thus, the result in this study provided evidence to support the speculation that CFB would be more effective in high complexity tasks (Doherty and Balzer 1988; Balzer et al., 1989).

6.1.4 Combined Feedback

CombFB was found to be effective in improving judgment performance in both low and high complexity tasks. Although it was hypothesised that CombFB would lead to greater improvement in judgment performance in a task of high complexity than in a low complexity one, the difference was in the expected direction but was not significant. When compared to TPF, CombFB did not lead to a significantly greater improvement in judgment performance than TPF in either task.

6.2 Moderating Effects of Self-insight

There have been no previous studies investigating the moderating effect of self-insight on the effectiveness of CFB. It was hypothesised in this study that self-insight moderated the effect of CFB in improving judgment performance in both tasks of low and high complexity. The results only supported the hypothesis for a task of high complexity, but not for a low complexity one. It is suggested that receiving information on their own judgment weights and models would be less informative, on average, for subjects in a low complexity task than for their counterparts in a task of high complexity.

This study found that CFB led to a larger improvement in judgment performance in a task of high complexity for low self-insight subjects than for their high self-insight counterparts. This result might partly explain the previous findings in psychology studies that CFB was generally not effective in improving judgment performance. Most of the psychology studies used simple generic tasks involving a small number of cues where subjects would be expected to have a reasonably high level of self-insight into their judgment weights and models. Under these conditions, CFB would not be effective in improving judgment performance.

Most previous studies have found that CombFB was not better than TPF in improving judgment performance. Given the moderating effect of self-insight on the CFB, it would be expected that the CFB component of CombFB would have the effect of improving judgment performance, especially for low self-insight subjects. Therefore, CombFB is likely to lead to an improvement in judgment performance that is over and above that of TPF alone. The results of this study were in the expected direction of the hypothesis, but failed to achieve the 5% significance level. The results might be due to the fact that both CFB and CombFB led to improvements in the judgment performance of subjects with low self-insight (see Tables 5.4 and 5.21) and due to the relative small sample sizes (11 in the CFB and CombFB groups) in Task 1A, and (11 in the CFB group and 12 in the CombFB group) in Task 1B.

6.3 Transfer of Learning

This study found that subjects who practised a configural processing task of high complexity performed better in another task involving a similar configuration of cues than their counterparts who practised a low complexity task, which did not involve configural cues. My results also show that subjects who practised a high complexity task required combined feedback for a better transfer of knowledge (that is, better judgment performance in Task 2).

The transfer of learning received through feedback has basically been ignored in both the accounting and psychology literatures. The results of this study relating to the transfer of learning are important to the auditing and training professions. In an ever changing business environment, it is virtually impossible to provide training to auditors in every new development. To prepare and facilitate auditors to exercise good judgment when faced with new clients, new transactions, and new business environments, audit firms have to train their auditors using complex training materials and tasks, as well as providing auditors with CombFB. Providing auditors with TPF only in training may improve their learning in that particular task, however, they may not be able to transfer their knowledge to other tasks.

6.4 Additional Analysis

6.4.1 Different Feedback Conditions in Promoting Configural Cue Processing

I am not aware of any previous accounting or auditing studies investigating the effects of different feedback conditions on the improvement of configural cue processing. Recall that domain- and task-specific knowledge specifies that configural cue processing is required in making a judgment in Task 1A. The effectiveness of different feedback conditions in improving configural cue processing was assessed in this study. Across two blocks of trials, it was found that TPF, CFB, and CombFB led to significant improvement in the extent of configural cue processing, measured by the judgment weight attributed to the interaction term. The NFB and OFB groups did not promote the adoption of configural cue processing. The average increase in weight attributed to the interaction term was largest for the CombFB group (8.22%), and the respective increases for the TPF and CFB groups were 4.978% and 4.619%. Therefore, CombFB is the most effective feedback in promoting configural cue processing amongst the five feedback conditions.

We also assessed the number of subjects who exhibited above-criterion (4%) weighting of the interaction term. The result was similar to that discussed above: TPF, CFB, and CombFB led to significantly more subjects attributing above-criterion weight to the interaction term (i.e., they exhibited configural cue processing) across two blocks of trials. The increase in the number of subjects who exhibited configural cue processing for each of the TPF, CFB, and CombFB groups, was 10, 6, and 11, respectively (see Table 5.26 for details). Again, NFB and OFB did not lead to a significant increase in the number of subjects who attributed above-criterion weight to the interaction term. The above results indicate that CombFB is the most effective feedback in promoting configural cue processing, both in terms of weight attributed to the interaction term and the increase in the number of subjects who exhibited configural cue processing.

6.4.2 Other Lens Model Parameters

In addition to judgment achievement (r_a), the lens model also specifies other parameters affecting judgment performance. Judgment consistency (Rs), linear knowledge (G) and non-linear knowledge (C) of the task are important parameters affecting judgment performance.

6.4.2.1 Judgment Consistency (Rs)

In a high complexity task involving configural cue processing (Task 1A), TPF, CFB, and CombFB led to a significant improvement in Rs in the second block of trials. Comparison of improvement between the two blocks of trials across the five feedback conditions revealed that TPF and CombFB were significantly better than NFB and OFB. There was no significant difference between TFB, CFB, and CombFB in improving Rs. The auditor subjects generally applied their judgment models fairly consistently in the two blocks of trials in Task 1A (Rs = 0.82 and 0.85, respectively).

In a task of low complexity that does not involve configural cue processing (Task 1B), OFB, TPF, and CombFB led to a significant improvement in Rs between the two blocks of trials and CFB led to a marginal improvement between blocks. Comparing the improvement in Rs across the feedback conditions revealed that OFB, TPF and CombFB led to a significantly higher improvement in Rs than NFB. The Rs was again fairly high in the two blocks of trials in Task 1B (0.85 and 0.89, respectively).

The results showing Rs was generally high were consistent with previous accounting and auditing studies (e.g., Hirst and Luckett, 1992; Hirst et al., 1999). Hirst et al., (1999) attributed the high Rs to "...the fact that subjects had reasonably well defined judgment strategies for the experiment task..." (p.297). This explanation can be applied in this study as the auditor subjects were familiar with making risk assessments regarding the possible misstatement of the accounts receivable balance. Virtually all audit engagements require auditors to make such risk assessments.

The impact of the various types of feedback on Rs found in the present study is generally consistent with earlier research. One exception was the results of Task 1B (the task of low complexity) where OFB led to a significantly higher Rs than NFB. While this was consistent with some psychology studies that used simple generic tasks (e.g., Holzworth and Doherty 1976; Schmitt et al. 1977), the results were not consistent with those of Hirst and Luckett (1992), Hirst et al (1999) and Tuttle and Stocks (1998). One possible reason may be that subjects in this study had both a high level of knowledge and day-to-day experience with the task. Tuttle and Stocks (1998) used undergraduate students as subjects, while, Hirst et al. (1999) had audit seniors performing a less familiar task – predicting corporate failure.

6.4.2.2 Linear Knowledge of Task (G)

For Task 1A, the results relating to G were similar to those of Rs. TPF, CFB, and CombFB led to a significant improvement in G in the second block of trials. TPF, CFB and CombFB were significantly more effective in improving G than NFB and OFB in Task 1A. The results were consistent with those of Hirst and Luckett (1992), where TPF led to significantly higher improvements in G than NFB. They did not study the effects of CFB or CombFB in their research.

For a low complexity task (Task 1B), OFB, TPF and CombFB led to significant improvements in G in the second block of trials. CFB had a marginal effect in improving G. Further, OFB, TPF and CombFB led to significantly higher improvements in G than did NFB. CFB did not differ significantly from the other feedback conditions in improving G. Some previous studies also found that OFB (Hirst and Luckett, 1992; Hirst et al., 1999; Tuttle and Stocks 1998) and TPF (Hirst and Luckett, 1992) were significantly better than NFB in improving G. In contrast to this study, Hirst et al., (1999) found TFB was not significantly different from NFB in improving G. They attributed this to the fact that generic TPF might be of little benefit to experienced subjects. It should be noted that the TPF provided to the auditor subjects in the bankruptcy task in Hirst et al., (1999) only stated that all three accounting ratios were important in predicting bankruptcy and whether the ratios were positively or negatively related to the judgment. In this study, both the functional form and actual weight of each cue were presented as TPF information to the subjects. The results of this study support the speculation of Hirst et al. (1999), that generic TPF might be of little benefit to experienced subjects, but more specific TPF might be effective in improving experienced subjects' linear knowledge of a task (G).

6.4.2.3 Non-linear Knowledge of Task (C)

No accounting research has investigated the effect of feedback conditions on the knowledge of the non-linear component of the task (C) and only a few psychology studies investigated and reported results on this aspect. In this study, TPF, CFB, and CombFB led to significant improvements in C in the second block of trials in Task 1A. However, when comparing the improvement in C across the five feedback conditions, only CombFB led to a marginal improvement ($p < 0.1$) compared to NFB and OFB.

The results of the lens model parameters suggest that the significant improvement in judgment accuracy (MAE) and judgment achievement (r_a) in Task 1A were mainly due to significant improvements in Rs, G, and C under the TPF, CFB, and CombFB conditions. For Task 1B, the improvement in judgment performance was due to improvement in Rs and G after subjects received OFB, TPF, and CombFB. CFB was more effective in improving judgment consistency (Rs) and linear knowledge of the task (G) for a task of high complexity. It was less effective for a low complexity one. On the other hand, OFB was effective in improving judgment performance in a low complexity task through improved judgment consistency (Rs) and linear knowledge of the task (G), but not in a high complexity one.

CHAPTER 7

CONCLUSION, LIMITATIONS, AND AREAS FOR FUTURE RESEARCH

7.1 Introduction

This chapter reviews the main aims, results and contributions of this thesis. This is followed by a discussion of the limitations of the study and future research opportunities.

7.2 Aims of the Study

The first major aim of this study has been to examine the effects of five feedback conditions on auditors' judgment performance in two levels of task complexity. Specifically, the effects of no feedback, outcome feedback, task properties feedback, cognitive feedback, and combined feedback on auditors' judgment performance in two tasks of different complexity were investigated in a controlled experimental setting with practising Big-5 auditors as subjects. Guided by Bonner's (1994) model of audit task complexity, this research manipulated task complexity in terms of the interaction of cues. The task of high complexity was operationalised as involving configural cue processing while the task of low complexity did not involve configural cue processing.

The second major aim is to examine the impact of cognitive feedback and combined feedback (cognitive plus task properties) in an audit setting involving both a configural and non-configural task. These issues have not previously been addressed in an audit environment.

The third major aim of the study has been to examine the moderating effect of self-insight on the effectiveness of various feedback conditions, especially CFB and CombFB in improving judgment performance. It is hypothesised that CFB would be more effective for subjects with low self-insight. This study is the first to investigate the moderating effect of self-insight in feedback research.

The fourth major aim of this study has been to examine the transfer of knowledge obtained from one task to a subsequent task. This aspect of learning is important as the business environment is always changing and auditors are constantly required to conduct audits in new business environments with new transactions.

This study has two subsidiary aims. First, this study examines the effect of various feedback conditions on promoting the extent of configural cue processing if the domain- and task-specific knowledge prescribes that such information processing is appropriate. Second, this study assesses the effect of various feedback conditions in improving the lens model parameters, such as judgment consistency (Rs), knowledge of the linear component (G) and nonlinear component (C) of the task. Analysis of these lens model parameters can identify the sources of improvement in judgment performance for each feedback condition.

7.3 Contributions of the Thesis

This study is the first to investigate the effects of various feedback conditions on auditors' judgement performance across different levels of task complexity. The results outlined in chapters 5 and 6 showed that whilst some types of feedback (E.g. TPF) are useful across different levels of task complexity , the performance of others types of feedback (E.g. OFB and CFB) depends on the level of task complexity. Specifically, OFB was more effective in the low complexity task, whilst CFB was more effective in the high complexity task. This has direct practical implications for the type of feedback practitioners should provide. Also, from a theoretical perspective it helps to reconcile some of the conflicting findings in earlier psychological and accounting studies. The results of this study suggest that the inconsistent findings in earlier studies may be due to differences in the complexity of the experimental tasks.

In line with the predictions, task properties feedback was found to be effective in improving auditors' judgment performance for both levels of task complexity. This result is robust and found in most accounting and psychology studies. The results of this study provide further evidence and extend our knowledge of the effectiveness of TPF in improving judgment performance for a task of high complexity involving configural cue combination in an auditing context.

Cognitive feedback has not been studied in the audit literature and is not a form of feedback that is presently used in audit literature. However, this study shows that it does have potential to be an effective means of improving judgement performance on more complex tasks. While cognitive feedback has not been found to be very useful in the psychology research using simple generic tasks, the present study supports the suggestions in that literature that it is likely to be more useful in more complex tasks. Consistent with predictions and the results of previous psychology studies, combined feedback was found to be effective in improving judgment performance for tasks of both levels of complexity. Also in line with the results of previous studies, combined feedback did not lead to a significantly higher improvement in judgment performance than task properties feedback in both levels of task complexity.

Compared with TPF or CFB, it was found in this study that CombFB did not lead to significantly higher judgment performance in tasks of both high and low complexity. Although the improvement in judgment performance was in the expected direction, especially in a task of low complexity, the difference was not statistically significant. Therefore, I could not rule out the value of CombFB. Also, given that CombFB was the most effective in the transfer of learning, this again suggests the need for further research to establish the conditions under which it will provide an incremental contribution to performance.

This study is the first to examine the moderating effect of self-insight on the effectiveness of feedback on judgment performance. It was found that self-insight did moderate the effect of cognitive feedback in improving judgment performance for a task of high complexity, but not for a low complexity task. CFB led to significant improvement in judgment performance for a high complexity task for low self-insight subjects, compared to their high self-insight counterparts. Subjects with low self-insight gained additional knowledge of their own judgment policies when given the cognitive feedback. This result may partly explain the previous findings in psychology studies that cognitive feedback was generally not effective in improving judgment performance. Most psychology studies employ simple generic tasks involving a small number of cues where subjects would have a reasonably high self-insight into their judgment weights and models. In such scenarios, cognitive feedback would not be effective in improving judgment performance.

The study of the transfer of learning is new to the accounting and auditing literatures. It is generally agreed that the learning and training effect is highest when the learning can be transferred to tasks both of similar and different natures. In this study, subjects who practised a configural task (Task 1A) and received CombFB had a significantly better performance in Task 2 (a configural case) when compared with their counterparts who received NFB or OFB. The performance on task 2 (a configural case) of those subjects in all five feedback conditions who practiced a configural task (Task 1A) was higher than for those subjects who practised a non- configural task (Task 1B) in all five feedback conditions. The results support the general findings in the learning literature that practice on a difficult task leads to higher learning and a transfer of knowledge effect (e.g., Goodman 1998; Salmon and Perkins 1989).

While there is general recognition in the audit judgement literature of the importance of configural cue processing (R. Ashton 1974; Brown and Solomon 1990, 1991) there has been very limited research on how to increase the level of configural cue processing. This study found that different feedback conditions had a differential effect in improving configural cue processing. The results of this study provide evidence that TPF, CFB, and CombFB can significantly improve auditors' learning of configural cue processing.

7.4 Limitations of This Study

7.4.1 Timing of Providing Feedback

In addition to the usual limitations of experimental studies, there are a few issues that are specific to this study. The first limitation is related to the timing of the provision of the various feedbacks. OFB was provided to subjects after each trial in the first block, whilst TPF, CFB, and CombFB were provided to subjects after the first block of trials. The timing of providing feedback is consistent with most feedback studies in the accounting and psychology literatures. Given that OFB was provided during the first block of trials, subjects in the OFB group might have improved their judgment (consistency) all the way through the experiment. This could potentially put the OFB group at a disadvantage when comparing the judgment performance between the two blocks of trials, as subjects' judgment performance in the first block might be higher than if there was no feedback information provided to them. However, this issue is unlikely to be a major problem in this study. In Task 1A, the improvement in judgment performance between the two blocks of trials for OFB was very similar to that of the NFB group (see Tables 5.4 and 5.7). One-way ANOVAs were conducted comparing the judgment accuracy and achievement of the five feedback conditions for Task 1A in the first block, and found no significant difference ($F_{(4, 106)} = 0.547$, $p = 0.701$) for MAE, and ($F_{(4, 106)} = 0.463$, $p = 0.763$) for r_a. Also previous research has found that it took some trials before OFB would improve judgment performance (e.g., Hirst and Luckett 1992).

The One-way ANOVAs for Task 1B had similar non-significant results ($F_{(4, 108)} = 0.379$, $p = 0.824$) for MAE and ($F_{(4, 108)} = 0.282$, $p = 0.889$) for ra. There was no significant difference in judgment performance in the first block of trials of Task 1B. For Task 1B, the OFB group exhibited significant improvement in the second block of trials, compared with that of the first block (see Tables 5.5 and 5.8) despite the possibility that providing OFB in the first block of trials might increase their judgment performance, making further improvement in the second block more difficult. Therefore, the provision of OFB in the first block appears to have had very little effect, if any, in either task.

7.4.2 Solicit of Subjective Weights

The second limitation is related to the first one, discussed above. OFB was provided in the first block of trials and the subjective weights of each subject were solicited after the first block. As such, subjects in the OFB group might be in a more advantageous position in reporting their judgment weights more accurately than other feedback conditions, as the latter received the relevant feedback information after their subjective weights were solicited. This timing of soliciting subjective weights was necessary to ensure that the information from the task properties feedback, etc. did not influence the subjective weights. Also, CFB and CombFB could not be provided in the first block of trials as they were derived from subjects' responses in the first block of trials. In order to assess whether the self-insight of subjects in the OFB was significantly different from other feedback conditions, one-way ANOVAs were conducted to compare the self-insight of subjects across the five feedback conditions which were all taken after block 1. There was no significant difference in self-insight across the feedback conditions for both tasks ($F_{(4, 106)} = 0.928$, $p = 0.450$ for Task 1A; $F_{(4, 108)} = 0.240$, $p = 0.915$ for Task 1B). The ANOVA results suggested that the timing of providing OFB did not significantly affect the subjective weights solicitation.

Another issue relating to solicitation of subjective weights needs to be considered. Subjects responded with cue weights for the previous full block of 16 cases. As they were providing their judgments, they were becoming increasingly familiar with the materials. It is possible that the weights supplied during the first 8 cases would be different to the second 8 cases. Further research could consider eliciting this information more frequently.

7.4.3 Definition of Self-insight

The third limitation relates to the definition of self-insight. In task 1A, two of the five cues interact and the pattern of interaction of these two cues affects the risk judgment in each trial. That is, the risk judgment in each trial is affected by six pieces of information – the five cues and the interaction of the two specific cues.

However, it would be very difficult for subjects to externalise the weights they assigned to the interaction of cues. Discussion with the audit partners, who participated in setting the benchmarks confirmed this. Therefore, I decided to solicit only the subjective weights of the individual cues from each subject. The self-insight of each subject was computed by correlating the main effects of the five cues of each subject's ANOVA model in the first block of trials with their respective subjective weights. As such, the self-insight computed for each subject might not be a perfect measure of the underlying construct as the interaction term was not included in the computation. As the risk judgments in Task 1B only require a linear combination of cues, the above discussion does not apply to this task. Given the problem of measurement, the self-insight correlation measure used was binary and other alternative measures could be considered.

7.4.4 Possible Overstatement of Self-insight

One of the concerns of using a factorial design is the possibility of an unrepresentative proportion of extreme cases in each block of trials that may have led to an overstatement of self-insight. For example, an auditor would seldom assess the risk of misstatement of an account balance if the majority or all of the major control procedures pertinent to that balance were not completed. Another type of extreme case that could lead to overstatement of self-insight, occurs when all the pertinent audit procedures were completed with no exceptions noted. In these two types of extreme cases, the judgments are relatively easy to make and subjects are likely to exhibit higher self-insight. However, it should be noted that this study compares the relative self-insight across groups. As such, assuming that the extreme cases would lead to the same extent of overstatement of self-sight across different feedback conditions, this factor is unlikely to affect the comparisons.

7.4.5 Number of Cases

I decided to have 32 trials of Task 1A and Task 1B in this experiment to avoid the possible loss of subjects' concentration and interest. Any smaller number of cases would cause difficulty wit the statistical analysis. The 32 cases were divided into two blocks to allow a comparison of the judgment performance of subjects across two blocks under various feedback conditions. To the extent that the different types of feedback may take longer to have an impact, the performance improvements could be understated.

7.4.6 Recruitment of Subjects

Most of the subjects who participated in this study were contacted from a list of graduates of a Hong Kong university. Therefore, there is the potential for them to be systematically different from auditors who graduated from other universities. However, this is considered unlikely. First, they attended the training programs provided by their respective CPA firm. They acquired more declarative knowledge from these training programs and this may mitigate such systematic differences, if any. Second, the auditors also acquired procedural knowledge through the actual conduct of audits. This further mitigates such potential systematic differences.

7.4.7 Number of Cues

Use of a factorial experimental design generally places a limit on the number of cues that can be used and the number of cue levels that can be manipulated. In this study, there are five cues in each of the experimental tasks and each cue can be at one of two levels. In practice, auditors may make such risk assessments based on a larger set of audit evidence.

7.4.8 Problems with Factorial Design

The use of a factorial design results in the use of cases that are not entirely representative of the task ecology (Cooksey 1996, p. 116; Trotman 1990). For example, the case involving five cues not being completed at the time of the judgment task is unlikely. In addition to the problem related to self-insight noted earlier, the inclusion of these extreme cases which are not representative of task ecology can impact consensus, accuracy, and confidence measures.

7.4.9 Statistical Analysis

A number of statistical issues need to be considered. First, use of a between-subjects design for the types of feedback within task conditions essentially assumes a single judgment model for all subjects for Task 1A and Task 1B. Any actual differences in cognitive processing will add to the estimation error. Second, in this thesis I selected a range of lens model statistics that were considered most relevant to testing the hypotheses. There is a huge range of potential lens model statistics that could be employed to further explain these results. Third, in addition to statistical significance, various measures of behavioural significance could also be considered

7.5 Areas for Future Studies

7.5.1 Measurement of Self-insight

The measurement of self-insight in tasks involving configural cue combinations could be improved and refined in future research so that the underlying construct can be more accurately measured. One option is to measure self-insight using the recognition approach (Reilly and Doherty 1992). In this approach, subjects are presented with the judgment policies of all the subjects in the study and are then asked to identify their own. This may be a useful approach to measure the self-insight of subjects, especially in a task involving complex and multiple configural cue combinations.

7.5.2 Process of Integration of Cues

In this study, I only considered the input and output of auditors' judgment processes under various feedback conditions. Auditors' judgment processes relating to the integration of cues was not investigated. I am not aware of any study investigating the effect of various feedback conditions in influencing subjects' processes of integrating information cues. This issue is particularly important if the task involves configural cue combinations as this will demand higher cognitive resources. Payne et al. (1988) found that subjects tended to adopt sub-optimal information integration rules or heuristics to perform a task that demanded a high level of cognitive resources. The provision of different types of feedback may have a differential effect in reducing the cognitive resource demand for such tasks, and as a result, may influence the approach and processes involved in integrating information cues. The results of this line of research would contribute to our knowledge of subjects' information processing techniques to improve judgement performance under different feedback conditions. This knowledge would assist in determining at what stage of the judgment process various types of feedback are most effective. Future research that examines the form of cognitive processing for different types of tasks would be useful as an input to the above research issues.

7.5.3 Combinations of Feedback

In this study, I only examined the effect of one combination of feedback – TPF and CFB (CombFB). Other combinations of feedback were not studied because of the limited number of subjects available. For example, psychology studies have investigated the combined effect of OFB and TPF (Hammond et al. 1973), OFB and CFB (Hammond and D. Summers 1972), and OFB and CombFB (Todd and Hammond 1965). Studying the effect of more combinations of feedback would add to our knowledge and help identify the most efficient and effective combination of feedback condition(s) to improve auditors' judgment performance.

7.5.4 Task Complexity

Two levels of task complexity were manipulated in the present study. Future studies may investigate the effect of various feedback conditions on auditors' judgment performance in tasks of a wider range of complexity. The results in these studies would provide us with a more comprehensive understanding of the effectiveness of different feedback conditions.

Further, different dimensions of task complexity in Bonner's (1994) model can be manipulated to generate tasks of different complexity. For example, information processing was found to be a limiting factor in auditors' judgment processes when information load (an information input factor) was high (Simnett and Trotman 1989). Providing different types of feedback may reduce the cognitive resource demand in either the information load (input complexity), process complexity, or output complexity. In this study, I only manipulated one aspect of the processing complexity – the functional form of processing.

7.5.5 Moderation Effect of Self-insight

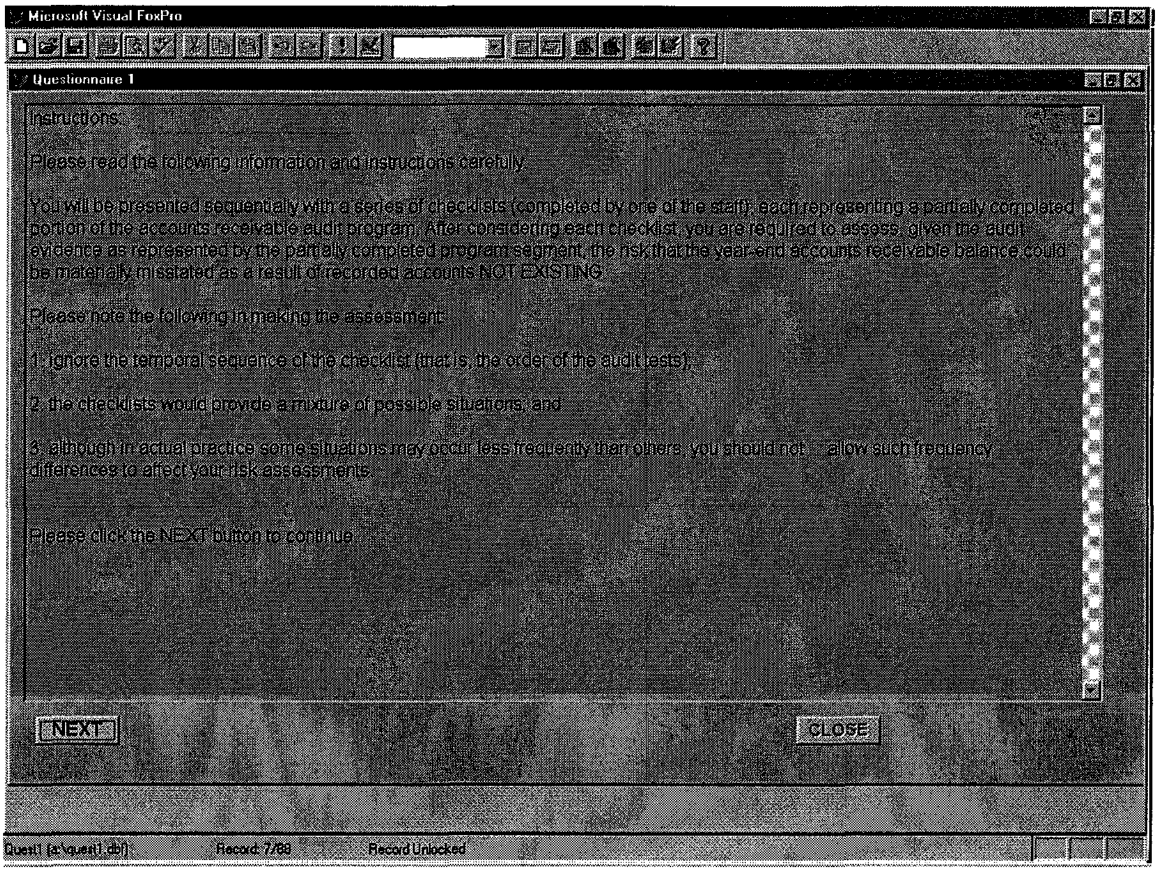
It was found in this study that self-insight moderated the effect of CFB in improving the judgment performance of auditors, especially in the task of high complexity. However, the effect was not significant for the low complexity task. It is important to identify the levels of self-insight that would trigger the moderating effect of CFB and combinations of feedback involving CFB, such as CombFB. The results have both theoretical and practical implications. Theoretically, the results may reconcile the inconsistent results of previous research relating to CFB and CombFB. The knowledge would also assist in the design of more effective training approaches provided to individual auditors.

7.5.6 Effectiveness of Improvement Over Time

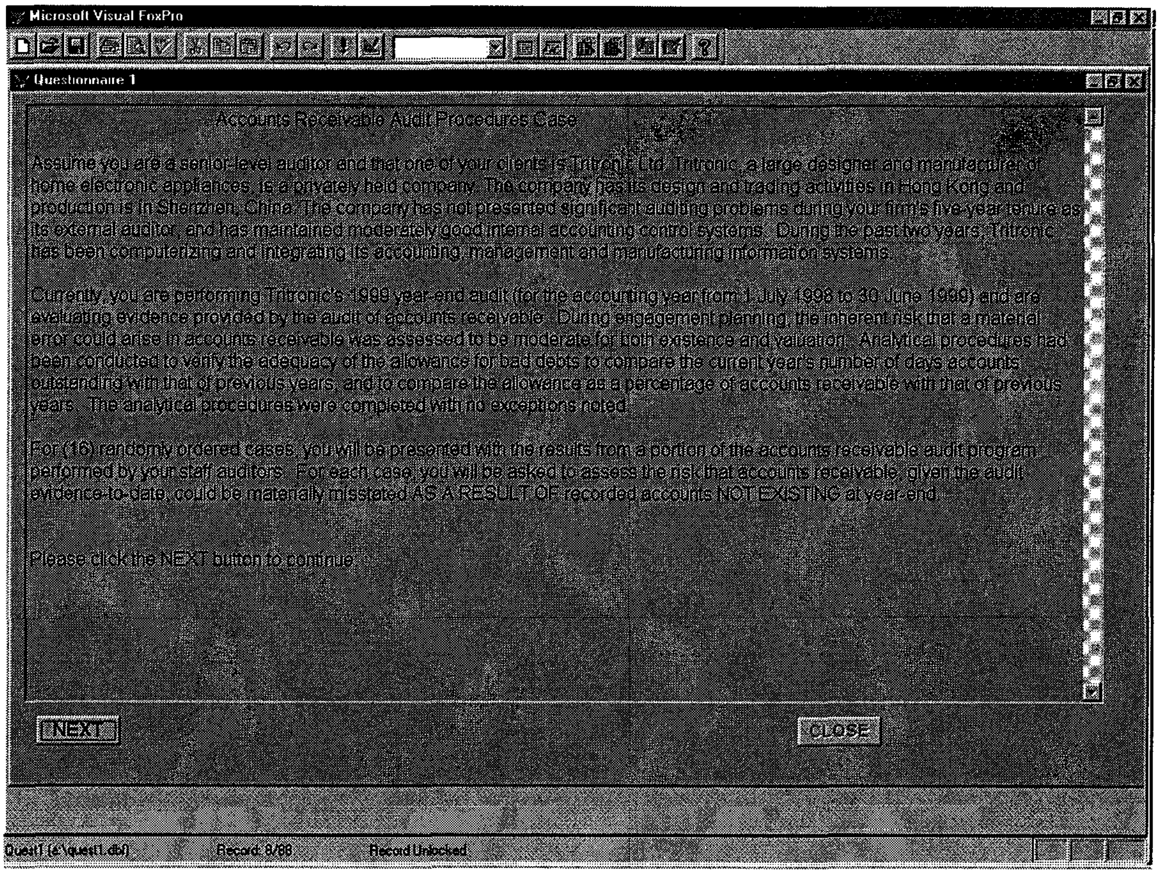
Consistent with previous research studying the effectiveness of different feedback conditions, judgment performance was measured over a short period of time. However, effective learning should last for a long period of time (Christina and Bjork 1991). Therefore, it is important to investigate the effect of different feedback conditions in improving judgment performance over a longer time period. This line of research may require the adoption of a longitudinal research methodology. However, considerable care in the design will be necessary to avoid internal validity threats such as maturity and experimental mortality.

Appendix A

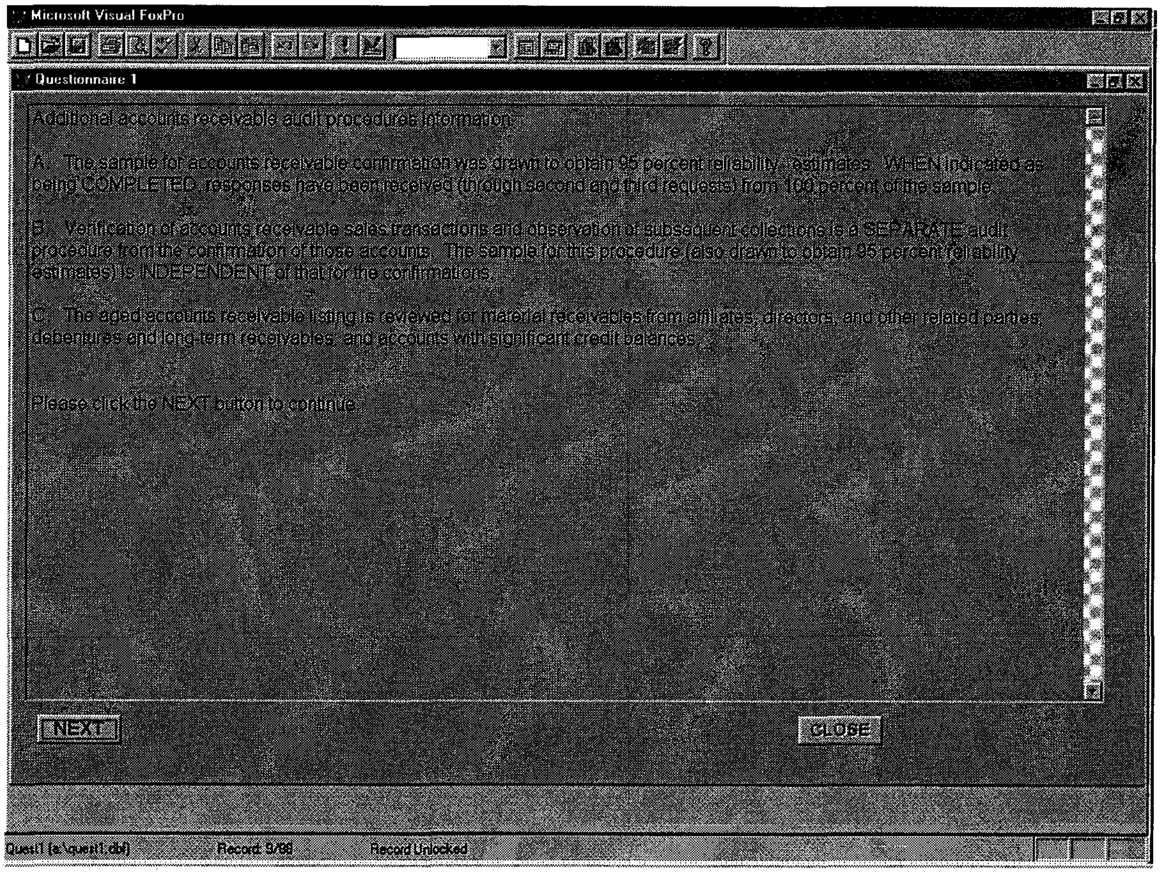
Instructions and Background Information of the First Block of 16 Trials for Task 1A



Appendix A (continued)

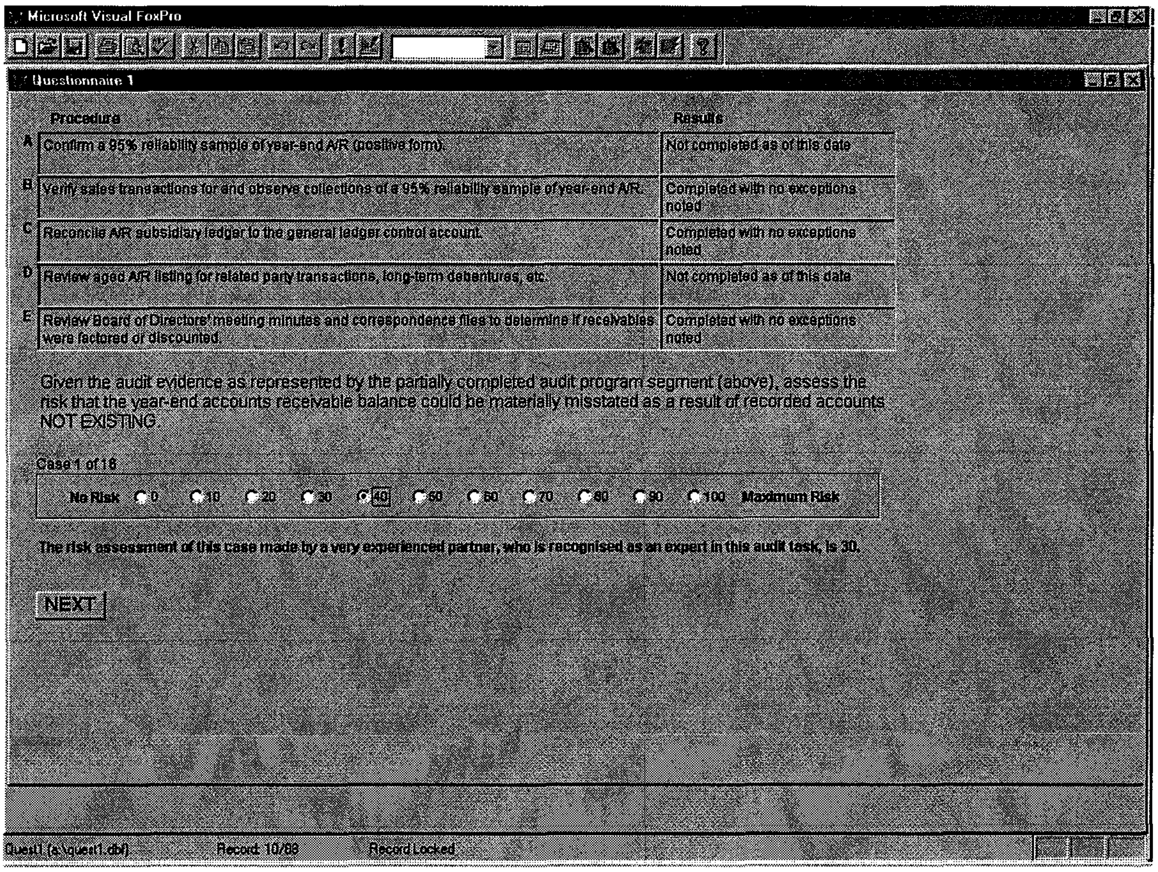


Appendix A (continued)



Appendix B

A Sample of Outcome Feedback Information for Task 1A



Note: The outcome feedback information shown below the response scale was only provided to subjects in the outcome feedback group. The information was shown on the screen once a subject clicked his/her risk assessment on the response scale. Subjects in other feedback groups would not receive such information.

REFERENCES

Abdolmohammadi, M., Wright, A., An Examination of the Effects of Experience and Task Complexity on Audit Judgments, *The Accounting Review*, (Vol.62, 1987), pp. 1-13.

Adelman, L., The Influence of Formal, Substantive and Contextual Task Properties on the Relative Effectiveness of Different Forms of Feedback in Multiple-cue Probability Learning Tasks, *Organizational Behavior and Human Decision Processes* (Vol.27, 1981), pp. 423-442.

Ashton, A.H., Does Consensus Imply Accuracy in Accounting Studies of Decision Making, *The Accounting Review*, (Vol.60(1), 1985), pp. 173-185.

Ashton, R.H., An Experimental Study of Internal Control Judgments, *Journal of Accounting Research* (Vol.12 (1), 1974), pp. 143-157.

-----, A Descriptive Study of Information Evaluation, *Journal of Accounting Research* (Vol.19 (1), 1981) pp. 42-61.

-----, Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback, and Justification, *Journal of Accounting Research* (Vol.28, Supplement 1990) pp. 148-180.

-----, and Ashton, A.H. (eds.*), Judgment and Decision Making Research in Accounting and Auditing* (1995), NY: Cambridge University Press.

-----, and Brown, P.R., Descriptive Modeling of Auditors' Internal Control Judgments: Replication and Extension, *Journal of Accounting Research*, (Vol.18 (1), 1980), pp. 269-277.

-----, and Kramer, S.S., Students as Surrogates in Behavioral Accounting Research: Some Evidence, *Journal of Accounting Research*, (Vol.18 (1), 1980), pp. 1-15.

Balke, W.M., Hammond, K.R., and Meyer, G.D., An Alternative Approach to Labor-Management Relations, *Administrative Science Quarterly*, (Vol.18, 1973), pp. 311-327.

Balzer, W.K., Doherty, M.E., and O'Connor Jr., R., Effects of Cognitive Feedback on Performance, *Psychological Bulletin* (Vol.106 (3), 1989), pp. 410-433.

-----, Sulsky, L.M., Hammer, L.B., Sumner, K.E., Birchenough, T.R., Martens, S.P., and Raymark, P.H., Task Information, Cognitive Information, or Functional Validity Information: Which Components of Cognitive Feedback Affect Performance?, *Organizational Behavior and Human Decision Processes* (Vol.53, 1992), pp. 35-54.

-----, Hammer, L.B., Sumner, K.E., Birchenough, T.R., Martens, S.P., and Raymark, P.H., Effects of Cognitive Feedback Components, Display Format, and Elaboration on Performance, *Organizational Behavior and Human Decision Processes* (Vol. 58, 1994), pp. 369-385.

Beach, L.R., and Mitchell, T.R., A Contingency Model for the Selection of Decision Strategies, *Academy of Management Review*, (Vol.3 (7), 1978), pp. 439-449).

Bonner, S.E., A Model of the Effects of Audit Task Complexity, Accounting, *Organizations and Society* (Vol.19 (3), 1994), pp. 213-234.

-----, and Pennington, N., Cognitive Processes and Knowledge as Determinants of Auditor Expertise, *Journal of Accounting Literature*, (Vol.10, 1991), pp. 1-50.

-----, and Walker, P.L., The Effects of Instruction and Experience on the Acquisition of Auditing Knowledge, *Accounting Review* (Vol.69 (1), 1994), pp. 157-178.

Bottomley, P.A., Doyle, J.R., and Green, R.H., Testing the Reliability of Weight Elicitation Methods: Direct Rating Versus Point Allocation, *Journal of Marketing Research*, (Vol.37, 2000), pp. 508-513.

Brehmer, B., Cognitive Dependence on Additive and Configural Cue-Criterion Relations, *American Journal of Psychology*, (Vol.82 (4), 1969), pp. 490-503.

-----, In One Word: Not from Experience, *Acta Psychologica* (Vol.73, 1980), pp. 223-241.

Bridges, D., *The Construction and Organisation of Knowledge in the University Curriculum*, paper presented in the British Educational Research Association Annual Conference, Belfast, 1998.

Brown, C.E., and Solomon, I., Auditor Configural Information Processing in Control Risk Assessment, *Auditing: A Journal of Practice and Theory* (Vol.9 (3),1990), pp. 17-38.

-----, and -----, Configural Information Processing in Auditing: The Role of Domain-Specific Knowledge, *The Accounting Review* (Vol.66 (1), 1991), pp. 100-119)

Carmerer, C., General Conditions for the Success of Boot-strapping Models, *Organizational Behavior and Human Performance*, (Vol.27(3), 1981), pp. 411-422.

Castellan, N.J. Jr., and Swaine, M., Long-Term Feedback and Differential Feedback Effects in Nonmetric Multiple-Cue Probability Learning, Behavioral Science, (Vol.22, 1977), pp. 116-128.

Christina, R.W., and Bjork, R.A., Optimizing Long-Term Retention and Transfer, in Druchman, D., and Bjork, R.A. (eds.), *In the Mind's Eye: Enhancing Human Performance*, Washington, DC: National Academy Press, 1991, pp. 23-56.

Clover, W., *Cognitive Feedback in the Selection Interview: Applying Social Judgment Theory in the Field*, Unpublished Ph.D. Dissertation, Bowling Green State University, 1979.

Colbert, J.L., Inherent Risk: An Investigation of Auditors' Judgments, *Accounting, Organizations and Society*, (Vol.13 (2), 1988), pp. 111-121.

-----, The Effect of Experience on Auditors' Judgments, *Journal of Accounting Literature*, (Vol.8, 1989), pp. 137-149.

Cook, D., The Effect of Frequency of Feedback on Attitudes and Performance, *Empirical Research in Accounting: Selected Studies* (1968) pp. 213-224.

Cooksey, R.W., *Judgment Analysis: Theory, Methods, and Applications*, (1996), San Diego, Cal.: Academic Press, Inc.

De Corte, E., Learning Theory and Instructional Science, in *Towards an Interdisciplinary Learning Science*, (Reimann, P., and Spada, H., editors), (1996) Oxford: Elsevier Science Ltd.

Deane, D.H., Feedforward and Multiple-Cue Probability Learning, (1970), *Institute of Behavioral Science, Cognitive Process Program, Report No. 137*, University of Colorado.

Deane, D.H., Hammond, K.R., and Summers, D.A., Acquisition and Application of Knowledge in Complex Inference Tasks, *Journal of Experimental Psychology*, (Vol.92 (1), 1972), pp. 20-26.

Doherty, M.E., and Balzer, W.K., Cognitive Feedback, in Brehmer, B., and Joyce, C.R.B. (ed.) *Human Judgment: The SJT View*, (1988), Amsterdam: Elsevier Science Publishers B.V., pp. 163-198.

Doyle, J.R., Green, R.H., and Bottomley, P.A., Judging Relative Importance: Direct Rating and Point Allocation Are Not Equivalent, *Organizational Behavior and Human Decision Processes*, (Vol.70 (1), 1997), pp. 65-72.

Earley, C.E., Knowledge Acquisition in Auditing: Training Novice Auditors to Recognize Cue Relationships in Real Estate Valuation, *The Accounting Review* (Vol.76 (1), 2001), pp. 81-97.

Einhorn, H.J., and Hogarth, R.M., Behavioral Decision Theory: Processes of Judgment and Choice, *Journal of Accounting Research*, (Vol.19 (1), 1981), pp. 1-31.

Fero, D.D., *A Lens Model Analysis of the Effects of Amount of Information and Mechanical Decision Making Aid on Clinical Judgment and Confidence*, Unpublished Ph.D. Dissertation, Bowling Green State University, 1975.

Frederick, D.M., and Libby, R., Expertise and Auditors' Judgment of Conjunctive Events, *Journal of Accounting Research*, (Vol.24 (2), 1986), pp. 270-290.

Galbraith, J.T., *Training Assessment Center Assessors: Applying Principles of Human Judgment*, Unpublished Ph.D. Dissertation, Bowling Green State University, 1984.

Gaumitz, B.R., Nunamaker, T.R., Surdick, J.J., .and Thomas, M.F., Auditor Consensus in Internal Control Evaluation and Audit Program Planning, *Journal of Accounting Research*, (Vol.20(2 Pt.II), 1982), pp. 745-755.

Goldberg, L.R., Simple Models or Simple Judgments, *American Psychologist*, (Vol.23, 1968), pp. 483-496.

Goodman, J.S., The Interactive Effects of Task and External Feedback on Practice Performance and Learning, *Organizational Behavior and Human Decision Processes* (Vol.76 (3), 1998), pp. 223-252.

Haccoun, R., Transfer and Retention: Let's Do Both and Avoid Dilemmas, *Applied Psychology: An International Review*, (Vol.46 (4), 1997), pp. 340-344.

Hamilton, R.E., and Wright, W.F., Internal Control Judgments and Effects of Experience: Replications and Extensions, *Journal of Accounting Research*, (Vol.20 (2 Pt. II), 1982), pp. 756-765.

Hammond, K.R., Computer Graphics as an Aid to Learning, *Science*, (Vol.172, 1971), pp. 903-908.

-----, *Reducing Disputes Among Experts*, Technical Report No. 268, Center for Research on Judgment and Policy, University of Colorado, 1987.

-----, *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injus*tice, NY: Oxford University Press, 1996.

-----, and Boyle, J.R., Quasi-Rational Quarrels, and New Conceptions of Feedback, *Bulletin of the British Psychological Society*, (Vol.24, 1971), pp. 103-113.

-----, Mumpower, J.L., and Smith, T.H., Linking Environmental Models with Models of Human Judgment: A Symmetrical Decision Aid, *IEEE Transactions on Systems, Man, and Cybernetics*, (Vol.7, 1977), pp. 358-367.

-----, and Boyle, P.J.R., Quasi-Rationality, Quarrels and New Conceptions of Feedback, *Bulletin of British Psychological Society*, (Vol.24, 1971), pp. 103-113).

-----, Hursch, C.J., and Todd, F.J., Analyzing the Components of Clinical Inference, *Psychological Review* (Vol.71, 1964) pp. 438-456.

-----, and Summers, Cognitive Control, *Psychological Review* (Vol.79, 1972) pp. 58-67.

-----, Summers, D.A., and Deane, D.H., Negative Effects of Outcome Feedback in Multiple-Cue Probability Learning, *Organizational Behavior and Human Performance*, (Vol.9, 1973), pp. 30-34.

Harmon, J., and Rohrbough, J., Social Judgment Analysis and Small Group Decision Making: Cognitive Feedback Effects on Individual and Collective Performance, *Organizational Behavior and Human Decision Processes* (1990) pp. 34-54.

Harrell, A., The Decision-Making Behavior of Air Force Officers and the Management Control Process, *The Accounting Review* (October 1977) pp. 833-841.

Hirst, M.K., and Luckett, P.F., The Relative Effectiveness of Different Types of Feedback in Performance Evaluation, *Behavioral Research in Accounting*, (Vol.4, 1992), pp. 1-22.

-----, -----, and Trotman, K.T., Effects of Feedback and Task Predictability on Task Learning and Judgment Accuracy, *Abacus*, (Vol.35 (3), 1999), pp. 286-301.

Hoffman, P.J., Earle, T.C., and Slovic, P., Multidimensional Functional Learning (MFL) and Some New Conceptions of Feedback, *Organizational Behavior and Human Performance*, (Vol.27, 1981), pp. 75-102.

Hogarth, R., *Judgment and Choice*, 2nd Edition, NY: John Wiley and Sons, 1987.

Holzworth, R.J., and Doherty, M.E., Feedback Effects in a Metric Multiple-Cue Probability Learning Task, *Bulletin of the Psychonomic Society*, (Vol.8 (1), 1976), pp. 1-3.

Hooper, C., and Trotman, K., Configural Cue Processing in Auditing: Further Evidence, *Accounting and Business Research*, (Vol. 26(2), 1996), pp. 125-136.

Hoskin, R. E., Opportunity Cost and Behavior, *Journal of Accounting Research* (1983) pp. 78-95.

Jiambalvo, J., Watson, D., and Baumler, J.V., An Examination of Performance Evaluation Decisions in CPA Firm Subunits, *Accounting, Organizations and Society*, (Vol.8 (1), 1983), pp. 13-29.

Joyce, E.J., Expert Judgment in Audit Program Planning, *Journal of Accounting Research* (Vol.14, Supplement 1976), pp. 29-60.

Kessler, L., and Ashton, R.H., Feedback and Prediction Achievement in Financial Analysis, *Journal of Accounting Research,* (Vol.19 (1), 1981) pp.146-162.

Kluger, V.N. and DeNisi, A., The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory, *Psychological Bulletin* (1996), Vol. 119 (2), pp. 254-284.

Kopelman, R.E., Objective Feedback, in Locke, E.A. (ed.), *Generalizing from Laboratory to Field Settings,* (1984), Lexington: Lexington Books.

Libby, R., *Accounting and Human Information Processing: Theory and Applications*, NJ: Prentice-Hall, 1981.

-----, Availability and the Generation of Hypotheses in Analytical Review, *Journal of Accounting Research*, (Vol.23 (2), 1985), pp. 646-665.

-----, and Libby, P.A., Expert Measurement and Mechanical Combination in Control Reliance Decisions, *The Accounting Review*, (Vol.64 (4), 1989), pp. 729-747.

-----, and Luft, J., Determinants of Judgment Performance in Accounting Settings: Ability, Knowledge, Motivation, and Environment, *Accounting, Organizations and Society*, (Vol.18 (5), 1993), pp. 425-450.

Lindell, M.K., Cognitive and Outcome Feedback in Multiple Cue Probability Learning Tasks, *Journal of Experimental Psychology: Human Learning and Memory*, (Vol.2, 1976), pp. 739-745.

Luckett, P., and Hirst, M.K., The Impact of Feedback on Inter-Rater Agreement and Self Insight in Performance Evaluation Decisions, *Accounting, Organizations and Society* (1989) pp. 379-387.

McCarthy, M.T., Webb, J.M., and Hancock, T.E., Form of Feedback Effects on Verb Learning and Near-Transfer Tasks by Sixth Graders, *Contemporary Educational Psychology*, (Vol.20, 1995), pp. 140-150.

McKenzie, C.R.M., Underweighting Alternatives and Overconfidence, *Organizational Behavior and Human Decision Processes* (Vol.71 (2), 1997), pp. 141.160.

Mear, R., and Firth, M., Cue Usage and Self-insight of Financial Analyst, *The Accounting Review*, (Vol.62 (1), 1987), pp. 176-182.

Newton, J.R., Judgment and Feedback in a Quasi-Clinical Situation, *Journal of Personality and Social Psychology*, (Vol.1 (4), 1965), pp. 336-342.

Nisbett, R.E., and Wilson, T.D., Telling More Than We Can Know: Verbal Reports on Mental Processes, *Psychological Review*, (Vol.84 1977), pp. 231-259.

Nystedt, L., and Magnusson, D., Cue Relevance and Feedback in a Clinical Prediction Task, *Organizational Behavior and Human Performance*, (Vol.9, 1973), pp. 100-109.

Poses, R.M., Cebul, R.D., Wigton, R.S., and Collins, M., Feedback on Stimulated Cases to Improve Clinical Judgment, *Medical Decision Making*, (Vol.6, 1986), p. 274.

Reilly, B.A., and Doherty, M.E., The Assessment of Self-insight in Judgment Policies, Organizational Behavior and Human Decision Processes (1992), Vol. 53, pp. 285-309.

Remus, W., Lim, K., and O'Connor, M., An Experiment on the Impact of Forecasts on Decisions, Working Paper, 1998.

-----, O'Connor, M., and Griggs, K., Does Feedback Improve the Accuracy of Recurrent Judgmental Forecasts?, *Organizational Behavior and Human Decision Processes* (1996) pp. 22-30.

Renkl, A., Learning mathematics from worked-out examples: Analyzing and fostering self-explanations, *European Journal of Psychology of Education* (1999), Vol. 14 (4), pp. 477-488.

Rohrbough, J., Improving the Quality of Group Judgment: Social Judgment Analysis and the Delphi Technique, *Organizational Behavior and Human Decision Processes* (1979) pp. 73-79.

Salmoni, A.W., Schmidt, R.A., and Walter, C.B., Knowledge of Results and Motor Learning: A Review and Critical Reappraisal, *Psychological Bulletin*, (Vol.95, 1984), pp. 355-386.

Salomon, G., and Perkins, D.N., Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon, *Educational Psychologist*, (Vol. 24(2), 1989), pp. 113-142.

Schmidt, R.A., Frequent Augmented Feedback Can Degrade Learning: Evidence and Interpretation, in Stelmach, G.E., and Requin, J. (ed.), *Tutorials in Motor Neuroscience*, The Netherlands: Kliwer, 1991, pp. 59-75.

Schmidt, R.A., and Bjork, R.A., Common Principles in Three Paradigms Suggest New Concepts for Training, *Psychological Science* (Vol. 3(4) 1992), pp. 207-217.

Schmitt, N., Coyle, B.W., and King,L., Feedback and Task Predictability as Determinants of Performance in Multiple Cue Probability Learning Tasks, *Organizational Behavior and Human Performance* (August 1976) pp. 388-402.

-----, -----, and Saari, B., Types of Task Information Feedback in Multiple-Cue Probability Learning, *Organizational Behavior and Human Performance* (April 1977), pp. 316-328.

Simnett, R., *Information Selection and Information Processing by Auditors: Effect of Varying Task Complexity*, Unpublished Ph.D. Dissertation, University of New South Wales, 1992.

-----, The Effect of Information Selection, Information Processing and Task Complexity on Predictive Accuracy of Auditors, *Accounting, Organizations and Society* (Vol. 21 (7/8), 1996) pp. 699-719.

-----, and Trotman, K., Auditor Versus Model: Information Choice and Information Processing, *The Accounting Review*, (Vol.64(3), 1989), pp. 514-528.

Simon, H.A., *The New Science of Management*, NY: Harper and Row, 1960.

Slovic, P., Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes, *Journal of Applied Psychology*, (Vol.53, 1969), pp. 255-263.

-----, Fleissner, D., and Bauman, W.S., Analyzing the Use of Information in Investment Decision Making: A Methodological Proposal, *Journal of Business*, (Vol.45, 1972), pp. 283-301.

-----, and Lichtenstein, S., Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment, *Organizational Behavior and Human Performance*, (Vol.6, 1971), pp. 649-744.

Stang, S.W., *An Interactive Judgment Analysis of Job Worth*, Unpublished Ph.D. Dissertation, Bowling Green State University, 1985.

Steinmann, D.O., Transfer of Lens Model Training, *Organizational Behavior and Human Performance* (Vol.12, 1974) pp. 1-16.

-----, The Effects of Cognitive Feedback and Task Complexity in Multiple -Cue Probability Learning, *Organizational Behavior and Human Performance* (Vol.15, 1976) pp. 168-179.

Steward, T.R., Judgment Analysis: Procedures, in Brehmer, B., and Joyce, C.R.B. (ed.) *Human Judgment: The SJT View*, (1988), Amsterdam: Elsevier Science Publishers B.V., pp. 41-74.

Summers, D.A., and Hammond, K.R., Inference Behavior in Multiple-cue Tasks Involving both Linear and Nonlinear Relations, *Journal of Experimental Psychology* (Vol.71, 1966) pp. 751-757.

-----, Taliaferro, J.D., and Fletcher, D.J., Judgment Policy and Interpersonal Learning, *Behavioral Science*, (Vol.15, 1970), pp. 514-521.

Summers, S.A., Summers, R.C., and Karkau, V.T., Judgments Based on Different Functional Relationships between Interacting Cues and a Criterion, *American Journal of Psychology* (Vol.82 (2), 1969), pp. 203-211.

Thompson, L., and DeHarpport, T., Social Judgment, Feedback, and Interpersonal learning, *Organizational Behavior and Human Decision Processes* (1994) pp. 327-345.

Todd, F.J, and Hammond, K.R., Differential Feedback in Two Multiple-cue Probability Learning Tasks, *Behavioral Science*, (Vol. 10, 1965), pp. 429-435.

Trotman, K.T., The Effect of the Distribution of Cases on Judgment Consensus, *Accounting and Finance*, (Vol. 30 (1) 1990), pp. 39-50.

-----, *Research Methods for Judgment and Decision Making Studies in Auditing*, Coopers and Lybrand Accounting Research Monograph No.3, Vic: Coopers and Lybrand Australia and AAANZ, 1996.

-----, Yetton, P.W., and Zimmer, I.R., Individual and Group Judgments of Internal Control Systems, *Journal of Accounting Research* (Vol.21 (1), 1983), pp. 286-292.

Tuttle, B., and Stocks, M.H., The Effects of Task Information and Outcome Feedback on Individuals' Insight into Their Decision Models, *Decision Science* (Vol.28(2), 1997), pp. 421-442.

-----, and -----, The Use of Outcome Feedback and Task Property Information by Subjects with Accounting-Domain Knowledge to Predict Financial Distress, *Behavioral Research in Accounting* (Vol.10, 1998), pp. 76-108.

Tucker, L.R., A Suggested Alternative Formulation in the Development of Hursch, Hammond and Hursch, and by Hammond, Husrch and Todd, *Psychological Review* (Vol.71, 1964), pp. 528-530.

Tuttle, B., and Stocks, M.H., The Effects Task Information and Outcome Feedback on Individuals' Insight into Their Decision Models, *Decision Science* (Spring 1997), Vol. 28(2), pp. 421-442.

-----, and -----., The Use of Outcome Feedback and Task Property Information by Subjects with Accounting – Domain Knowledge to Predict Financial Distress, *Behavioral Research in Accounting* (1998), Vol. 10, pp. 76-108.

Wood, R.E., Task Complexity: Definition of the Construct, *Organizational Behavior and Human Decision Processes*, (Vol.37 (1), 1986), pp. 60-82.

Wright, A., The Impact of Prior Working Papers on Auditor Evidential Planning Judgments, *Accounting, Organizations and Society*, (Vol.13 (6), 1988), pp. 595-605.

Wright, W.F., Self-insight into the Cognitive Process of Financial Information, *Accounting, Organizations and Society*, (Vol.2 (4), 1977), pp. 323-331.

-----, Audit Judgment Consensus and Experience, in Ferris, K.R. (ed.), *Behavioral Accounting Research: A Critical Analysis*, 1988, Ohio: Century VII Publishing Company.

Zhu, S-H., and Anderson, N.H., Self-Estimation of Weight Parameter in Multiattribute Analysis, *Organizational Behavior and Human Decision Processes*, (Vol.48, 1991), pp. 36-54.