Design Issues in Audit JDM Experiments

Ken T. Trotman*
University of New South Wales, Australia

This paper outlines the purposes of audit judgment and decision making (audit JDM) research and explains the basic principles of experiments and experimental designs. Choices of dependent and independent variables are discussed, together with choices of manipulating variables within and between subjects. The use and effectiveness of randomization to control for extraneous variables is discussed together with various internal validity threats that can exist. Finally, four design choices affecting internal validity are addressed. These design issues relate to the use of control groups, manipulation checks, controlled experiments and incentives. Many of the choices raised in this paper are discussed in more detail in subsequent papers in this issue of the journal.

Key words: audit judgment, decision making, design choices, experiments, research methods.

INTRODUCTION

The importance of judgment in auditing has been recognised for many decades (e.g. AICPA, 1955; Mautz, 1959) and today pervades most auditing standards. It is therefore not surprising that a large literature examining audit judgments has developed. This research is generally labelled as Judgment and Decision Making (JDM) Audit Research (also called ‘behavioral decision making research in auditing’). Over the last 25 years this research has been one of the dominant research paradigms in auditing. In addition, this research has had a major impact on audit practice as illustrated in the AICPA monograph ‘Auditing Practice, Research and Education: A Productive Collaboration’ (Bell and Wright, 1995).

The purpose of this paper is to outline some of the design choices that need to be made in conducting audit judgment research and some of the pitfalls in making these choices. The remainder of the paper is structured as follows. Section 2 outlines the purposes of audit judgment research and Section 3 considers experimental design including the choice of independent and dependent variables. Sections 4 and 5 consider control of extraneous variables and other design choices.

PURPOSE OF AUDIT JUDGMENT RESEARCH

The basic aim of this research is to improve auditor judgments. Before improvements can be suggested and empirically tested it is necessary to evaluate the quality of different types of judgments, understand how judgments are made and understand what are the key factors that affect auditor performance under different conditions. Early research starting in the 1970’s (e.g. Ashton, 1974) on auditor judgments had the purpose of evaluating auditor judgment quality...
and discovering how the judgments are made. Specific questions addressed included: What is the level of consensus on auditor judgments? How accurate are auditor judgments? Are auditor judgments consistent across time? Do auditors use simplifying judgmental rules called heuristics and do these lead to significant biases in auditor judgments? What information do auditors use in making judgments and what pieces of information do they place most reliance on?

The purpose of most audit judgment since the mid 1980's has been to examine the determinants of judgment performance. Performance is a function of ability, knowledge, environment and motivation (Einhorn and Hogarth, 1981; Libby, 1983). The earlier research discussed above concentrated on ability (measuring performance, documenting judgment strategies and identifying information processing errors). Over the last 15 years much of the research has addressed the other three determinants of performance, particularly knowledge. This research studies how information is encoded, stored and retrieved from memory and how these knowledge structure differences are related to performance differences (Libby and Luft, 1993). The following types of questions are addressed: What knowledge is needed to complete different audit tasks? When, how and how well the knowledge is acquired? What are the cognitive processes through which the knowledge is brought to bear on the decision tasks? What is the role of the auditors’ task specific knowledge in problem recognition, hypothesis generation and information search? What factors affect auditor memory for audit evidence? How does this memory affect auditor judgments? (Libby, 1995). A recent trend has been to extend this work to consider differences between industry and non-industry specialists (Solomon et al, 1999; Low, 2001).

Much of the current audit JDM research has examined the interactions between the various factors affecting performance. For example, does the review process improve performance and, if so, what are the specific sources of process gain? What impact does prior involvement with a client have on subsequent audit judgments? What impact does accountability have on performance? (See Tan, 2001, in this issue of IJA, for details of these research methods.)

Answering the above questions provides insights for suggesting remedies for any deficiencies found since it is necessary to understand the decision process in order to improve it. For example, the motivation for research on knowledge is that if we can better understand the knowledge of an expert auditor and how it is acquired, it may be possible to develop training and decision aids which improve the performance of novices.

In examining judgment performance there has also been a trend towards understanding predecisional behaviour. For example, before auditors make judgments they often need to search for and select the appropriate information. In respect to analytical review, Koonce (1993) suggests there is a diagnostic, sequential and interactive process involving mental model representations, hypothesis generation, information search and hypothesis evaluation prior to the judgment being made. While earlier research looked at individual aspects of this process more recent research has looked at the relationship between these stages and the impact on the final judgment (e.g. Asare and Wright, 2000). Asare and Wright (2001), in this issue of IJA, provide a detailed overview of research on hypothesis generation/evaluation and the research methods used. Bedard and Biggs (1991) provide an example of the use of protocol analysis (verbal reports) in an analytical review study.

EXPERIMENTS AND EXPERIMENTAL DESIGNS

The most common method of examining auditor judgments is the use of an experiment. An experiment is defined by Kerlinger (1973) as ‘a scientific investigation in which an investigator manipulates and controls one or more independent variables and observes the dependent variable or variables for variation concomitant to the manipulation of the independent variables. An experimental design, then, is one in which the investigator manipulates at least one independent variable’.

Consistent with psychology experiments, most audit judgment studies manipulate one or more variables and all other factors are held constant or measured. The variables manipulated by the researcher are called independent variables. The experiments examine the effect of the manipulations on another variable called the
dependent variable. For example, if a researcher was examining the impact of a number of types of decision aids on going concern judgments, the different types of decision aids would be the independent variable and the going concern judgment would be the dependent variable. In many audit judgment experiments, the various treatment groups consist of different levels/amounts of the independent variable. For some experiments there will be an experimental group (a group that receives the manipulation) and a control group (a group that does not receive the manipulation). The experimentalist’s comparative advantage lies in the ability to abstract fundamental components of real world settings and control other potentially influential variables. This allows the researcher to look at certain variables while holding others constant. Thus the researcher can disentangle interrelated elements of the accounting/ auditing setting which affect behaviour and test the effects of conditions which do not yet exist in that setting (Libby and Luft, 1993).

Kerlinger (1973) defines a research design as ‘the flow, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance’. Thus research designs have two basic purposes: to provide answers to research questions and to control error variance. The first factor is discussed below. The second factor refers to the control of extraneous independent variables and is discussed in Section 3.

Research designs are developed to allow the researcher to answer research questions as validly, objectively, accurately and economically as possible. Research questions are expressed in terms of hypotheses which are derived from theory. Theories in auditing are developed based on previous research (including descriptive research) and/or theories from other disciplines such as economics, cognitive psychology and social psychology. These theories ‘borrowed’ from other disciplines are adapted based on a knowledge of the auditing process (e.g. Gibbins, 1984). For example, research on auditors’ knowledge and memory and how this affects auditor judgments is based heavily on theory developed from cognitive psychology. Much of the literature on audit groups is based on theories from the small group literature in social psychology. However, this theory needs to be adapted to take account of the unique attributes of auditing including the sequential and hierarchical nature of the review process, the loss functions faced by auditors, and prior involvement with clients. Gibbins and Swieringa (1995) suggest that while general judgment research often depends on theory and is not particularly dependent on problem settings, audit judgment studies are both theory driven and setting sensitive. They note that audit judgment research focuses on incentives, constraints, tasks, structures, and other characteristics of applied settings as potential modifiers or determinants of human judgment processes. Gibbins (2001) elaborates on the importance of context.

Once you have decided on your research question, decisions have to be made on the specific independent variables to be manipulated and the dependent variables to be observed. Examples of independent variables used in auditing studies include the presence or absence of various internal controls (Ashton, 1974; Joyce, 1976); accounts receivable checklists (completed or not completed) (Brown and Solomon, 1991); analytical review result (Brown and Solomon, 1990; Cohen and Kida, 1989); base rate of fraud (Joyce and Biddle, 1981b); various types of group formats (Trotman and Yetton, 1985; Trotman, 1985); the type of inherited hypothesis presented as their superior’s most likely estimate of the cause of fluctuations in ratios (Libby, 1985); task complexity – unstructured, semi-structured or structured tasks (Abdolmohammadi and Wright, 1987); engagement risk (Hackenbrack and Nelson, 1996); amount of information (Simnett, 1996); hypothesis frame – viable or fail (Kida, 1984); the presence or absence of decision aid (Kachelmeier and Messier, 1990); time pressure (McDaniel, 1990; Glover, 1997); type of task and experience level (Bonner, 1990); source credibility (Bamber, 1983; Hirst, 1994); whether the explanation for the hypothesised non-error cause is written prior to or after an elicitation of the probability (Koonce, 1992); review anticipation (Tan, 1995; Koonce et al, 1995); client integrity (Peecher, 1995); role of the auditor (preparer or reviewer) (Libby and Trotman, 1993); accountability (Kennedy, 1993; Hoffman and Patton, 1997; Tan and Kao, 1999); incentives (Libby and Lipe, 1992; Brown et al, 1999); hypothesis set size (Bhattacharjee et al, 1999); probability of loss (Nelson and Kinney, 1997).
Manipulation of the above independent variables generally involves the presence or absence of some cue/treatment (e.g. existence or non-existence of an internal control, presence or absence of feedback, decision aid, group decision making, etc.), amount of a variable (e.g. base rate of fraud, number of pieces of information provided, number of alternative explanations), timing of receiving a variable (e.g. before or after making an initial assessment), or type of variable presented (e.g. type of group format; task complexity – structured, semi-structured, unstructured; hypotheses frame – viable or failure; source of information; and strength of information). These variables are manipulated by the contents of the research instrument given to different treatment groups and sometimes a control group. In addition, some independent variables are called subject or measured variables. In this case, subjects are selected (or at least put into different treatment groups) based on the amount or type of a particular measured variable. In audit studies, this variable is most often experience or a specific type of knowledge such as technical knowledge (Kennedy and Peecher, 1997; Tan and Libby, 1997) or industry knowledge (Solomon et al., 1999). Some early audit judgment studies measured personality variables such as intolerance to ambiguity (e.g. see Gul, 1984) and decision style (Driver and Mock, 1975).

In designing a study it is important to consider the construct validity of the independent variable. The term refers to confounding, which means that what one researcher interprets as a causal relationship between theoretical constructs labelled A and B could be interpreted by another researcher as a causal relationship between constructs A and Y, X and B, or X and Y (Cook and Campbell, 1979). For example, in early studies examining experience effects, experience was measured by the number of years a person had been with an audit firm. Later studies, however, showed the need to relate experience to tasks (e.g. Abdolmohammadi and Wright, 1987; Bonner, 1990) and the distinction between general experience and expertise in the performance of specific tasks (e.g. Bonner and Lewis, 1990; Davis and Solomon, 1989; Libby and Luft, 1993).

Choices have to be made about the number of independent variables to be included in the design and the number of levels of each independent variable. An increase in either will substantially increase the number of subjects required for the study. The number of independent variables is generally determined by the research question. For example, assume the research question you are interested in is the impact of a specific decision aid on the judgments of audit seniors and audit managers. In this case the presence/absence of a decision aid would be manipulated and the senior/manager variable would be measured giving a 2 x 2 design. However, if you believe that the impact will depend on how complex the task is (which could be varied across two, three, four, etc. levels), the level of training (again could be varied over a number of levels) and whether the participants are industry specialists or not, the addition of these variables will increase the number of treatment groups substantially (and therefore the number of subjects). If any treatment variable is manipulated across more than two levels the number of subjects required again increases. Given the difficulty in obtaining sufficient participants all of the above variables cannot be addressed in the one study. For example, industry specialist and training could be held constant, e.g. only use industry specialists and provide training to all participants. Which variables to hold constant and which to manipulate will depend on your theoretical developments. For example, your theoretical development may predict that managers do behave differently to seniors on more complex tasks because of the specific way in which they store the information in memory. If data is collected on how auditors store this information in memory (by recall or recognition tests) this allows the researcher to not only see if the hypothesised effects exist (e.g. managers outperform seniors on complex tasks using decision aids) but also why they do so.

Many new researchers manipulate too many variables and then have a very small number of subjects per cell. Consequently the chances of finding any significant results are very small. My suggestion is to design your study to incorporate all variables you really would like to include. Then having found out how many subjects you can obtain for the experiment, determine what is the maximum number of cells you can have in the experiment and still have enough power to get significant results. At that stage you have to make some very difficult
choices about what independent variables to drop. Again the importance of theory should be emphasised here. It is important to consider the theoretical support for each variable and the links between variables. 

Choices also need to be made on whether each independent variable should be manipulated within subjects (subjects get all levels of a treatment) or between subjects (subjects get only one level of a treatment). For example, consider the situation where the researcher is testing the effect of decision aids on auditors’ going concern judgments for cases of different complexity (complex/easy). There are four cells (2 x 2): decision aid/complex cases, decision aid/easy cases, no decision aid/complex cases, and no decision aid/easy cases. Using a between-subjects design, each subject would be randomly allocated to only one of the four treatments. Using a within-subjects design, all subjects would be exposed to all four treatments. It is also possible to combine between and within-subject designs in one design where one independent variable has a different group of subjects for each level whereas all subjects are exposed to each level of the other independent variable. In the above example, it would be possible to have task complexity as a within-subjects factor and decision aid as a between-subjects factor.

There are a number of advantages of using a within-subjects design. First, they require considerably less subjects to conduct the experiment. Second, a within-subjects design results in greater statistical power because subjects act as their own control in comparison among treatments effects. In a between-subjects design, differences between subjects can constitute a substantial source of variance. In the within-subjects design, subject variables (e.g. experience) are constant over the experiment and therefore it is more sensitive to treatment effects. Third, they are very effective in examining certain types of research questions. For example, learning effects and cue usage.

However, there are a number of disadvantages of using within-subjects designs. First, the concern is that when subjects receive a number of treatments, they are more likely to be sensitised as to what the hypotheses are. This can lead to demand effects which refers to characteristics of a situation that are likely to result in subjects behaving in a particular way that is not connected to differences in treatments. For example, subjects generally want to impress the researcher, to illustrate their competence, etc. and if they guess the hypotheses of the researcher they may make judgments which they believe the researcher desires or expects. Second, the repeated testing of subjects in a within-subjects design can lead to practice effects. This practice effect can be positive, resulting in subjects doing better as the experiment progresses, or can be negative due to factors such as fatigue and boredom. Third, carry-over effects occur when the effect of one treatment persists in some form at the point of measurement of another treatment. For example, a strategy learned in one treatment of a within-subjects design could be carried over and used in another treatment. This situation could exist if a researcher was comparing the effects of outcome feedback with task properties feedback. It is likely that subjects would develop a strategy based on whichever feedback was received first, and this strategy could be carried over to when judgments are made using the second strategy. It should be noted that in both the psychological (e.g. Birnbaum, 1999) and accounting literatures (see Schepanski et al., 1992) there is debate about when within and between subject designs are appropriate. Thus for each variable to be manipulated, the researcher needs to consider whether to manipulate the variable within or between based on the factors discussed above.

The choice of a dependent variable is critical. This choice depends on a number of factors, including its relationship to the independent variables. If manipulation of the independent variables results in an effect, it is important that the dependent variable picks up the direction and magnitude of that effect. That is, it must be sensitive to changes in the independent variables. Examples of dependent variables used in auditing studies include the strength of the internal control system (Ashton, 1974); time required for various audit procedures (Joyce, 1976; Mock and Wright, 1993); risk that an account balance is materially misstated (Brown and Solomon, 1991); probability estimates (Joyce and Biddle, 1981a, 1981b); number of plausible, implausible and high frequency hypotheses generated (Libby and Frederick, 1990); elicitation of an initial probability (Koonce, 1992; Asare and Wright, 1995); prediction of company failure.
(Kida, 1980, 1984; Simnett and Trotman, 1989); auditor’s likelihood assessment (Heiman, 1990; Peecher, 1995); determination of a non-statistical sample size (Kachelmeier and Messier, 1990); memory for different types of evidence (Choo and Trotman, 1991; Tan, 1995); most likely EPS to be reported (Libby and Kinney, 2000).

In summary, experiments are of particular value in examining audit issues particularly given the importance of judgment in auditing. Unlike many other areas of accounting research, databases to examine auditor judgments are not generally available. Examining audit firm workpapers is a possibility where access is available (see Mock and Turner, 2001, in this issue), but these judgments in the workpapers are affected by many factors and it is difficult to establish the effect of particular factors. This can be achieved by the use of experiments which have two major benefits. The first relates to the strength with which a causal relationship can be inferred. Second, they have the advantage of allowing the researcher to precisely manipulate one or more variables which are of interest to the researcher. However, for an experiment to be successful, extraneous factors must be controlled as discussed in the following section.

CONTROLLING EXTRANEOUS VARIABLES

Experiments are internally valid when the variation in the dependent variable can be unambiguously attributed to the manipulation of the independent variable(s) (Campbell and Stanley, 1963). Therefore, for an experiment to be internally valid it is necessary to ensure that the observed effect is only caused by the independent variable(s) and is not influenced by other extraneous variables. For example, assume a researcher is interested in the effect of a new decision aid on auditor going concern judgments. They split subjects between a control group (no decision aid) and a treatment group (decision aid). If the treatment group has more general audit experience, more experience with going concern clients or training using the decision aid, these factors could confound the results of the experiment. That is, experience and training are extraneous variables.

The main method of controlling extraneous variables is randomization of subjects between treatment groups. One aim of randomization is to ensure that both known and unknown extraneous variables do not systematically bias the study results. Randomization is the only method of controlling for all possible extraneous variables. If properly carried out then the different treatment groups can be considered statistically equal in all possible ways (Kerlinger, 1964).

While randomization is very effective in ruling out most threats to internal validity, there are some it does not rule out. Two such examples are ‘imitation of treatments’ and ‘resentful demoralization of respondents receiving less desirable treatments’ (Cook and Campbell, 1979). First, imitation of treatments can occur when subjects in one treatment group learn the information intended for another treatment group. In a controlled experiment examining auditor judgments, this is usually prevented by asking (and observing) that subjects work independently. When questionnaires are mailed to participants additional care is needed to ensure that subjects from different treatment groups do not communicate. Second, if participants from one treatment (or the control group) become aware that another group of subjects is in a more favourable treatment (e.g. more or better information) they may become less motivated. For example, if subjects were asked to predict corporate failure for a group of companies and one treatment group was given cash flow information and the control group was not, subjects in the control group may feel that they are at a disadvantage compared to the treatment group if they become aware of the information the treatment group has. For example, Simnett and Trotman (1989), in studying the effect of information choice and information processing on auditors’ going concern judgments, allowed one group to select a set of ratios they wished to use, while the other group were given ratios selected by the researchers. We believed that if all the subjects completed the experiment in the one room, the subjects who were not permitted to select their ratios may have been less motivated if they believed they were at a disadvantage compared to the other group. Consequently, we administered the two treatments in different rooms with subjects on arrival being randomly allocated to one of the rooms. Similar procedures are often desirable in studies comparing types of feedback, decision aids and group formats.
Another internal validity threat that can occur in a randomized experiment is differential mortality between treatment groups. While treatment groups may initially be equivalent because of randomization, more subjects may drop out of one treatment than the other. This is more of a problem if the subjects that drop out have a common trait (e.g. less motivation than the average subject). This can result in the remaining subjects in one treatment group not being comparable to those in another group. This is more likely to occur when questionnaires are mailed to participants rather than a controlled experiment. For example, if there is a high information load group and a low information load group, less of the high information group may respond because the task takes longer. In varying the amount of information between treatments as above, the researcher also needs to be aware of another threat to internal validity, namely ‘maturation’. This refers to any naturally occurring processes within subjects which could affect performance (Cook and Campbell, 1979). For example, in designing certain audit judgment studies (e.g. repetitive within subject designs) it is important to be aware of potential learning, fatigue and boredom effects. These effects affect internal validity if the maturation rate varies between treatments. For example, in comparing high and low information load groups, fatigue may occur much earlier for the high information load group.

In conducting audit experiments it is often necessary to collect data over a number of weeks at various staff training venues. Where possible each training session should be treated as a separate block, so that treatments are randomly allocated within each block. This blocked randomization can occur for most independent variables but not for certain measured variables. For example, staff training for seniors may occur in one week and for managers a few weeks later. In this case, the researcher should provide an instruction to subjects to please not discuss the nature of the experiment for a certain period and to check that no event occurred in the environment during that period that could affect results (e.g. major corporate collapse may impact going concern judgments).

In summary, while randomization is critical to eliminating many potential internal validity threats, care is still needed to address the above internal validity threats. It is also necessary to control for subject variables, control for experimenter bias and use counterbalancing in within-subjects designs to control for order effects.

Levels of expertise and/or particular types of experience is a very common independent variable in audit judgment studies (see Libby and Luft, 1993 for a review). Obviously, subjects cannot be randomly allocated to these treatment groups and therefore these attributes need to be measured. For example, if a researcher is comparing the going concern judgments of audit managers and audit seniors it is important to ensure that other extraneous factors besides audit level are ruled out. For example, there may also be systematic differences in the industries they work in, the number of going concern decisions they have been involved in, changes in professional training over time, etc. In the above situation, where the researcher is dealing with quasi-experimental groups, the researcher cannot rely on randomization to rule out internal validity threats. Instead he/she needs to make all threats explicit and rule each one out (Cook and Campbell, 1979).

Internal validity for these types of studies is considerably strengthened by following the ‘expertise paradigm’ (see Libby and Luft, 1993). This paradigm suggests that three conditions need to be met. That is, three guidelines that experimental studies of knowledge issues should follow are: (a) it needs to be a conceptual perspective which specifies the knowledge and cognitive processes. This means that because different audit tasks require different amounts of knowledge, the hypotheses should be developed in advance about the effects of specific knowledge elements or their organisation on behaviour. The researcher needs to specify the knowledge necessary to complete a particular task as well as when, how, and how well it will be acquired; (b) it is necessary to demonstrate a hypothesised knowledge difference and/or its effects on performance by constructing an experimental task where the observable implications of using and not using knowledge are different. That is, if the task is such that two people with different levels of knowledge are still likely to come to the same judgment then it would be an inappropriate task; (c) the existence of a knowledge effect can best be established by manipulating stimuli
and/or context factors in comparing individuals with different experiences. An example of this would be to have participants complete two different types of tasks: one where you would expect no differences because the knowledge levels of participants would be approximately equal, and a different task where you would expect differences because of differences in knowledge levels. This third guideline is very powerful in that it has the ability to eliminate alternative explanations for knowledge or performance differences. For example, when comparing performance of managers and seniors, one potential reason why managers do better than seniors is a self selection process. There are less managers than seniors and probably only the best seniors go on to be managers. Therefore it can be argued that managers are generally better performers. However, to eliminate this alternative possibility the researcher should use two tasks, one where no differences are predicted and one where differences are predicted because of knowledge differences. If these results are found there is a much stronger case that the knowledge differences are the cause of differences in performance. Libby and Luft (1993) provide examples of a number of papers that have adopted these guidelines and describe how the guidelines are operationalized. Tan (2001) in this issue, further addresses design in measuring knowledge effects.

It is also necessary to control ‘experimenter bias’, which refers to the unintentional biasing effects a researcher can have on the results. Where possible the same person should administer the experiment at each sitting of the experiment. Written instructions should be given and the administrator of the experiment should work from the same script each time. Where more than one room is used at the same time in the experiment it is obviously necessary to have more than one person administering the experiment and it is important that they are as consistent as possible in the administration.

Another type of experimenter bias can relate to coding of data. This can occur if the responses are qualitative (for example, generation of hypotheses, items recalled, lists of errors, etc.). In this case it is important to have two coders who are blind to the hypotheses when coding and then examine the inter-rater reliability. Gibbins and Newton (1994) provide a description of well designed coding to eliminate experimental bias. Asare and Wright (2001), in this issue, provide further advice on coding.

In a within subjects design it is important to avoid order effects. For example, assume you are comparing the performance of experienced and inexperienced auditors on both a simple and complex task. If subjects always complete the simple task followed by the complex task and you find both groups perform better on the complex task, your finding is confounded with a possible learning effect. One solution is to use counterbalancing. In this case half the seniors and half the managers would do the simple task first and the other half would do the complex task first. When the number of treatments is large, partial counterbalancing is appropriate. One such method is randomized counterbalancing where the researcher gives each subject cues in a different random order. This is particularly important when the order of reading the cues can affect judgments.

In some cases a potential confounding effect cannot be avoided. In this case it is important for the confounding to work against your hypothesis. For example, assume you are designing a study to compare a specific aspect of performance between audit managers and audit seniors. Your hypothesis is that on this specific task managers will outperform seniors. Given the problems of obtaining the number of subjects, you decide to offer a small gift certificate to each participant. Should the amount be the same for managers and seniors or should managers be given more since each dollar is probably worth less to them given their higher salary compared to seniors. Alternative A is to pay the managers larger amounts. If the managers’ performance is better than the seniors, thus supporting your hypothesis, there are at least two potential reasons. Managers are better at this task (as predicted) or managers were more motivated because they received a larger gift voucher. Alternative B is to give them all the same dollar value gift certificate. The confounding here is that seniors may be more motivated by the gift certificate than managers and therefore work harder. If no difference was found between managers and seniors this could be because managers are not better at the task or because there were differences in motivation between managers and seniors. Given your initial belief that managers will outperform seniors, alterna-
tive B seems the preferred option because it works against your hypothesis.

DESIGN CHOICES AFFECTING INTERNAL VALIDITY

This section discusses four specific design choices related to control groups, manipulation checks, controlled experiments and incentives. First, a decision needs to be made on whether to include a control group in addition to treatment groups. Subjects in the control group are treated exactly the same as those in the experimental group except they do not receive the treatment. There are two main benefits of control groups: (a) only by having a control group can a researcher conclude whether or not the treatment produced results different from what would have occurred in the absence of the control group; (b) it controls for rival hypotheses because the extraneous variables are held constant (Christensen, 1994). The major disadvantage of a control group is that it requires additional subjects. Often the question you are addressing determines the need for a control group. For example, if you are interested in which of two feedback types improves auditor judgments most, then a control group is not necessary. However, if you wish to know how much a particular type of feedback improves judgments compared to no feedback then you need a control group.

Second, a decision has to be made on the type of manipulation checks to be included in the study. Earlier I noted the importance of the construct validity of the independent variable. In order to examine the validity of the manipulated variables, it is necessary to carry out a manipulation check by obtaining a measure of the independent variables. Manipulation checks can be used to check that relevant pieces of information that form the treatment are encoded. For example, if the factor being manipulated is accountability (the subjects will/will not have their work reviewed by a superior) it is important to know that subjects encoded this information. A question would be asked to ascertain whether the subject believed their work was to be reviewed or not. Manipulation checks are also used to check the strength and direction of cues perceived by the subject are consistent with the treatments. For example, in a belief revision study examining going concern judgments, subjects are given an equal number of positive and negative cues and it is important that the strength of the positive and negative cues is approximately equal. In this case the manipulation check would ask subjects to rate the direction and strength of the impact of each of the items on their likelihood a company would remain a going concern.

In situations where the manipulations are physically implemented there may be no need for a manipulation check, for example, when the order of the cues is varied between treatments. Similarly, for situations where the framing of the question is manipulated. For example, one group is asked the probability of a company remaining a going concern and the other treatment group asked the probability of the same company remaining viable.

A third design choice is whether to conduct a controlled experiment (experiment conducted in the presence of the researcher or his/her assistants) or a non-controlled experiment (distributed to subjects by researcher or audit firm and returned). Internal validity is considerably higher in a controlled experiment as the researcher has control over the time spent on the task, who completed the task, information consulted, distractions, subjects working independently, etc. Also manipulation checks can be more difficult to administer in a non-controlled setting because of the problem of sensitizing subjects to the manipulations and subjects returning to earlier parts of the questionnaire. However, given the increased difficulty of obtaining subjects for audit judgment research particularly with respect to access to staff training courses, it is inevitable that more non-controlled experiments will be carried out. The researcher needs to be aware of potential validity threats and with the increased use of computerized research instruments many of these threats can be reduced or eliminated.

Fourth, a decision also has to be made on whether any monetary incentives need to be paid to subjects. While this practice is very common in experimental economics studies and psychology studies involving students, it is less common in the audit JDM literature. Historically, audit firms have had a commitment to assisting this type of research and incentive payments have not been necessary. In these experimental settings, while the effects of participants’ incentives might be diminished, their
direction should not be altered and therefore the directional effects of treatments should not be altered (Libby et al, 2001). There are also certain things a researcher can do to enhance the motivation of participants. For example, providing realistic tasks, involving partners in the introduction to the experiment as evidence of firm commitment to the project, offering to provide feedback to participants and involving the researchers in data collection (I believe it is not a task to be delegated to research assistants). If it is decided that some form of monetary incentive needs to be provided, great care is needed to ensure that the monetary incentive does not result in subjects using non-typical behaviours. For example, a single prize may result in subjects adopting high-variance strategies which they would not normally do (Dopuch, 1992). My belief is that unless incentives is an independent variable I would either not pay incentives or give the same amount to all subjects.

CONCLUSION

This paper addresses a range of research design choices that need to be made in developing a JDM audit research study. It outlines the choices necessary in the selection of independent and dependent variables, alternatives in controlling extraneous variables and a range of design choices affecting internal validity. The comparative advantage of an experiment is it can manipulate a small number of variables while holding other variables constant. It is critical that the appropriate design choices are made to control for extraneous variables and thus enhance the internal validity of the experiment.

Many of the above design choices are discussed in greater detail in the following five articles of this journal. In the following paper Peccher and Solomon (2001) elaborate on a number of design issues addressed in this paper and in particular the critical role of theory and some important research traps to be avoided. The next two papers consider design issues in specific types of experiments. Asare and Wright (2001) consider analytical procedure experiments and Tan (2001) experiments related to knowledge effects. Gibbins (2001) emphasizes the importance of incorporating context into your designs as well as design issues related to questionnaire construction to obtain data through retrospective recall by auditors. Mock and Turner (2001) discuss design choices for archival research examining audit working papers.

ACKNOWLEDGEMENTS

The helpful comments of Mahreen Hasan, Ira Solomon and Arnie Wright are appreciated.

NOTES

1. Some of this discussion is based on Trotman (1996).

REFERENCES


AUTHOR PROFILE

Ken Trotman is a Scientia Professor in the School of Accounting, University of New South Wales. His main research interests relate to auditor judgments. He is presently associate editor of ‘Auditing: A Journal of Practice and Theory’. He has received a number of academic awards including the AAA Audit Section 2001 ‘Outstanding Educator Award’.