Metaknowledge in Higher Education: Self-Assessment Accuracy and its Association with Academic Achievement

TONY BLACKWOOD

Doctor of Business Administration

2010

A Study of Metaknowledge in Higher Education

TONY BLACKWOOD

A thesis submitted in partial fulfilment

of the requirements of the University of Northumbria

for the degree of Professional Doctorate

Research Undertaken in Newcastle Business School

January 2010

Abstract

An appreciation of the extent of one's knowledge is known as metaknowledge and it has been argued that students' ability to distinguish between what they know, and what they do not, is an important influence on academic success. However, previous research suggests a general tendency for individuals to display overconfidence in their knowledge, by overestimating how much they know. This study assessed the ability of learners studying business in higher education to appreciate the extent of their own knowledge and investigated the association between this capability and academic achievement. It therefore contributes towards answering broader questions regarding how well individuals are able to assess their own capabilities and what the implications of this are.

Quantitative methodology was employed and multiple-choice tests used to investigate how accurately students were able to assess the extent of their knowledge of issues addressed in their study programmes. Analysis of over 12,500 judgements provided by 508 respondents revealed a general tendency for overconfidence and indicated that this was greater for males, older participants and particularly, for Chinese students. Consequently, interventions designed to moderate overconfidence may be particularly valuable for these sub-groups. In terms of its potential implications for learning, the research indicated that better metaknowledge was positively associated with higher levels of academic performance, particularly for those in their first year studying at the university. Consequently, while metacognitive skills, such as accurate self-monitoring, are typically poorly addressed in business schools, the findings from this study suggest that initiatives to improve self monitoring accuracy may be effective in enhancing student learning. Additionally, such interventions have other potential benefits for learners, since metacognitive monitoring skills may also usefully inform lifestyle decisions, as well as improving the chances of success in business and may therefore be particularly beneficial for business students.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank many people for their support and encouragement during the time I have spent on the DBA programme and latterly while producing this thesis. Firstly, I would like to express my thanks to my wife Fiona and children Ashleigh and Ryan for their support and patience during the many hours I have spent working on this study.

I would also like to thank my supervision team, Andy Robson and Craig Mahoney for their guidance, valuable feedback, encouragement and general support. I am also grateful for the interest and support I have received from my parents, other family members, friends and colleagues. Finally, I would like to express my gratitude to the students who agreed to participate in the study and the following for their assistance in data collection:

Angela McGrane, Ian Charity, John Dickinson, Philip Shrives, Keith Anderson, David Grundy, Barrie Michinson, Martin Foo, Lawrence Ferry, Karen Boyd, Tom Baines and Bob Burroughs.

Tony Blackwood

(January 2010)

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work.

Name: Tony Blackwood

Signature:

Date:

1. Introduction

'To know that we know what we know and that we do not know what we do not know; that is true knowledge'

Confucius

1.1 General Aims

The purpose of this study is to contribute towards answering questions regarding how effectively individuals are able to assess their own capabilities and what the implications of this are. More specifically, its substantive aims are to assess the ability of learners studying business in higher education to appreciate the extent of their own knowledge and to investigate the association between this capability and academic achievement. The research is guided by a conceptual framework incorporating metacognition and investigates theory suggesting a general tendency for individuals to overestimate their own knowledge. The practical relevance of the findings is addressed by considering their implications for professional practitioners working in higher education. Pring (2000) argues that generalisable conclusions, however tentative, are more likely to influence policy makers in professional practice and the methodological aim of the study is to employ a quantitative approach to generate findings which may be generalised. A non-experimental, fixed research design is applied, using a questionnaire to gather data from students studying in higher education. This is analysed using statistical techniques to test specific hypotheses developed in the study and findings are discussed in the context of relevant academic literature.

1.2 The Significance of Metaknowledge

Educational research has been criticised on the grounds that it does not support professional practice sufficiently (Pring, 2000). It is therefore important to consider the significance of the issues addressed in this study and their potential implications for educational practice. Academic performance in education can be improved by self-regulated learning (Stone, 2000), an approach in which students take responsibility for their own learning (Paris and Winograd, 2003). It typically entails various activities which emphasise 'autonomy and control by the individual, who monitors, directs and regulates actions towards goals of information acquisition, expanding expertise and, self-improvement' (Paris and Paris 2001, p.89). Since it tends to enhance performance, self-regulated learning is of interest to researchers investigating how students may become more independent and effective learners (Paris and Paris, 2001). Azevedo (2009), for example, emphasises the increasing importance of research which investigates self-regulatory processes that contribute to learning and academic achievement and it has been argued that the driver of self-regulated learning is metacognition (Paris and Paris, 2001).

1.2.1 Metacognition

Metacognition is 'knowledge about cognition and cognitive processes' (Schraw, 2009 p.34) and is concerned with the appraisal and management of learning, through monitoring, evaluation and planning (Everson and Tobias, 1998). Thus, it is concerned with thinking about thinking, a process which has been referred to as 'metathinking' (Crittenden and Woodside, 2007 p.37). Metacognitive skills can empower students to be better managers of their own learning (Paris and Winograd, 1990) and recent research has addressed the need to assist students in developing a better understanding of their own cognition and thinking (Pintrich 2002). Montalvo and Torres (2004) suggest that to enhance learning, the education process should assist students in developing their awareness of their own thinking and in becoming their own teachers. However, Sternberg (1998) argues that, because teachers are either unaware of them or not sure how they should be taught, metacognitive skills are given insufficient attention in education. In the specific context of business education, Crittenden and Woodside

(2007) argue that, while students and business managers should appreciate the importance of metathinking, few business academic programmes incorporate it and academics in business schools have not done a good job in integrating it into classroom activities in a manner which will enhance students' appreciation of how metathinking may enhance success in business. This is despite the fact that learning programmes often emphasise the importance of critical thinking skills and one of the main reasons for its inadequate treatment, may be that it was recognised as an area for study only relatively recently (Crittenden and Woodside, 2007).

1.2.2 Self Monitoring

An important aspect of metacognition is monitoring, which relates to students' ability to assess their own learning and is one of the main features of self-regulated learning (Isaacson and Fujita, 2006). A characteristic of effective learners is having a realistic appreciation of their own strengths and weaknesses, since this helps to direct attention towards deficiencies (Boud and Falchikov, 1989). Thus, while one of the key purposes of education is to provide students with knowledge (Jehng, Johnson and Anderson, 1993), metacognitive theory emphasises the importance of learners also having a good appreciation of the extent of their own knowledge. This concept has been referred to as metaknowledge (Russo and Schoemaker, 1992) and the 20th century British philosopher Alfred North Whitehead emphasised its relevance when observing that, 'it is not ignorance, but ignorance of ignorance, that is the death of knowledge' (Dunning et al, 2003 p.86). A good appreciation of the extent of our knowledge is important because it helps us to understand whether we need more information (Renner and Renner, 2001) and in an educational context can therefore provide students with a basis for determining future learning strategies (Grimes, 2002), by directing attention to gaps in knowledge. Poor metaknowledge on the other hand means that a learner is not motivated to address these deficiencies (Sternberg, 1998). However, while good metaknowledge may be desirable, the most common research finding is that people tend to display overconfidence in their knowledge, by overestimating its accuracy (Renner and Renner, 2001). This tendency, has been described as 'a fundamental feature of human psychology' (Bar-Tal, Sarid and Kishon-Rabin, 2001 p.77) and has been reported in findings of research conducted in many settings. In higher education, for

example, Smith, Shields and Washburn (2003), report the tendency for undergraduates in higher education to over-optimistically assess when their knowledge is satisfactory. Previous studies have also reported that the less knowledgeable are particularly poor at monitoring their own knowledge, which represents a two-fold problem for these individuals. Not only do they not know, but they also lack the skills to appreciate this and unless they are made aware of it, will be unable to remedy this inadequacy (Dunning et al, 2003). Consequently, professional educators have a responsibility to assist students in appreciating how much they do not know (Kennedy, Lawton and Plumlee, 2002).

1.3 Implications for Professional Practice

1.3.1 The Transfer of Research into Professional Practice

As well as producing knowledge, the aim of educational research should be to improve practice (Winch, 2001) and Gersten at al (1997) argue that the researcher's role in knowledge production is irrelevant unless it has practical application. However, Hemsley-Brown and Sharp (2004) explain that while literature addressing the use of research in education has tended to suggest a need for better understanding of effective strategies for teaching and learning, numerous studies suggest that education professionals rarely use research findings to inform their practice. They argue that the way in which research is designed may restrict its impact on professional practice, since the goals of researchers and practitioners tend to differ in that the former may wish to develop new knowledge, while the latter require solutions to practical problems. A distinction can be drawn here between Mode 1 research, in which knowledge is produced with little consideration for its deployment in practice (Worrall, 2008) and Mode 2, which attempts to solve complex problems of relevance to professional practitioners (Van Aken, 2005). The former is typically disseminated via peer reviewed academic journals to an audience of fellow academics (usually those working in a narrow field of study) and may never be used to inform practice. Mode 2 research on the other hand is less exclusive and elitist and targets practitioners as well as academics, aiming to provide insights which may inform practice (Worrall, 2008).

To facilitate the transfer of research into professional practice, it is important that, as well as being rigorous, it also satisfies the needs of practitioners. However, educational research has also been criticised in recent years for addressing issues that are not their main concerns and which have little relevance in practice (Winch, 2001) and consequently, Aram and Salipante (2003) argue that research questions should relate to the context in which practitioners operate. Gersten et al (1997) highlight the problems associated with research findings indicating the value of interventions which were derived in settings which do not mirror the environments in which practitioners work, suggesting that while such studies may be informative as to the potential value of such initiatives for learning, they will not be implemented unless they can be adapted to reflect the settings and constraints under which practitioners operate. Research conducted by Zeuli (1994) also suggests that research is more likely to be considered by practitioners when it can be related to their personal experiences and addresses interventions with the potential for implementation in the classroom. Consequently, greater involvement of practitioners in conducting research may usefully assist researchers in appreciating the conditions in which they operate and therefore enhance its impact on professional practice (Huberman, 1990) and Hemsley-Brown and Sharp (2004) suggest involving them in the focus and design of research as well as subsequent follow up activities. Sparks (1988) also found that a key factor influencing professional educators' propensity to adopt new initiatives was the potential benefits for students. Thus research findings which suggest a particular intervention may be associated with higher levels of academic achievement may be more likely to impact on professional practice through implementation by practitioners.

Hemsley-Brown and Sharp (2004) suggest that another issue contributing to the limited impact of educational research on professional practice is the manner in which knowledge is developed in social research, which gives rise to findings being challenged on the grounds of the context of research, as well as the generalisability and validity of findings. As Gersten et al (1997 p.466) point out, education, unlike medicine for example, has no 'magic bullets', no chemical remedies to eliminate problems.

Hillage et al (1998) also argue that the lack of encouragement for academics to disseminate research findings to practitioners contributes to the limited impact of research on professional practice. Worrall (2008) suggests that academics may be more

highly motivated by enhancing their academic reputation than generating research of greater value to practitioners and that in an alternative value system, practitioner focused research would be held in as high esteem as that published in highly rated academic journals. Hemsley-Brown and Sharp (2004) highlight proposals addressing this issue, suggesting that academics should be rewarded for dissemination of research findings to practitioners, who should be targeted via alternative outlets for publication.

Gersten et al (1997) also highlight the problem of researchers underestimating the difficulties educational practitioners have in implementing initiatives addressed in research studies due to them being based on rather abstract theory. Consequently, research findings should be communicated in a manner which professional practitioners are able to interpret (Aram and Salipante, 2003). Hillage et al (1998) argue that professional practice is poorly informed by research, due to problems in the manner in which it is disseminated. These include the inaccessibility of academic journals to non-academic practitioners and the lack of time and support for practitioners in accessing research. In a review of literature addressing the use of research in education, Hemsley-Brown and Sharp (2004) therefore emphasise the value of making it more accessible to practitioners and reducing the use of academic jargon in encouraging the use of research findings in practice. However, in the context of this study it is worth noting that the practitioners for whom the research is relevant are professional educators working in higher education. For these, problems associated with the manner in which research is disseminated may be less important since they tend to have easier access to research findings and are likely to be more familiar with terminology used in academic articles.

In conclusion, while the impact of educational research on professional practice may be low, researchers can enhance this by better marketing of their knowledge to practitioners and this may be achieved through anticipating and meeting their needs, considering the value of participative research and disseminating findings clearly and effectively (Hemsley-Brown and Sharp, 2004).

1.3.2 Implications of This Study

The findings from this study may usefully inform professional educators responsible for the design, management and promotion of learning activities in higher education. For

example, a general tendency for poor self-assessment of knowledge would suggest the introduction of specific remedial interventions since, if people are typically overconfident, they can be assisted in developing debiasing techniques (Klayman et al, 1999). However, the potential impact of doing so will be influenced by the extent to which metaknowledge is associated with academic achievement and findings from this aspect of the research will therefore assist in assessing this. A strong association between knowledge monitoring accuracy and academic performance for example, while not necessarily implying causality, would provide evidence suggesting that initiatives to promote the development of metaknowledge may be effective in raising levels of academic achievement.

1.4 Context of the Study

The research setting for the study was Newcastle Business School, a large provider of business education based in the University of Northumbria, which delivers a range of business-related study programmes, from undergraduate up to doctoral level, to students from a wide variety of cultural backgrounds. The school's learning and teaching strategy emphasises the importance of independent learning and since this can be facilitated by accurate self-monitoring, it provides an appropriate setting for the investigation of metaknowledge and its association with academic performance.

1.5 The Author's Position and Interest

Issues for social research often arise from the researchers' desire to understand how their life relates to others' in society (Baker, 1999) and most researchers, particularly in doctoral studies, are motivated by personal interest in the subject (James and Vinnicombe, 2002). In respect of this study, a recent personal development activity in which the author participated provided insights in terms of personal interest in the concept of metaknowledge and how this may have been influenced by relationships with others. During this activity, feedback received from a psychometric tool evaluating personal preferences indicated a preference for others to be sure of their facts and suggested that those communicating with the author should not 'pretend that you

know more than you do'. This highlights a preference prompting personal interest in how able people are to appreciate how much they know, which has been further provoked by professional experiences. The author is an academic member of staff at the institution in which the investigation took place, with responsibility for delivering learning activities and managing study programmes. Previous experiences in that environment have suggested that, on occasions, learners tend to overestimate their knowledge. For example, in conversations with students who were new to the university, some claimed that through previous learning, they already had good knowledge of some of the topics which were included in their study programme. However, subsequently, it often became clear to the author, and occasionally to the students in question, that they were not as knowledgeable as they initially believed and still had much to learn about the issues in question. Additionally, during classroom activities students have occasionally made strong claims for knowledge which, through probing and further discussion, proved to be unwarranted.

The author's experiences in supervising research dissertations for students at undergraduate and postgraduate level have also been influential in stimulating interest in metaknowledge. Towards the end of their time spent working on a dissertation, some students have tended to display greater awareness of how much there was to know and a better appreciation of the extent of their own knowledge about the topic. However, this state was usually attained only after having spent a significant amount of time investigating the issue in question, which suggests that knowledge and accurate knowledge monitoring may be positively associated.

1.6 Issues Addressed to Achieve the Aims of the Research

Numerous issues were addressed in order to achieve the aims of the study. Initially, the theoretical framework in which the study is located was established and approaches adopted in previous research, as well as findings arising from those studies, considered. This assisted in developing specific research questions, to facilitate the achievement of the overall aim of assessing students' ability to appreciate the extent of their own knowledge and investigating the association between this capability and academic achievement. Research methodology and design were also addressed. Various

alternative approaches may be used in social research and therefore, in order to assist in its interpretation, it is necessary to explain the approach adopted and clarify the philosophical assumptions on which it is based. Since this study is quantitative, the advantages and potential limitations of this methodology in social research were addressed, as well as the manner in which it influences research design. The design employed to achieve the aims of the study were also established and this entailed considering alternative possible approaches and identifying one which was appropriate. Since quantitative studies entail operationalising the concept under investigation, an appropriate indicator of metaknowledge was established. The development of an appropriate method for collecting data was also considered and the specific setting in which this was to take place established. This included developing initiatives to enhance the reliability and validity of the research and, as the aim was to generate findings which may be generalised to the population under investigation, establishing an appropriate sampling strategy. Since data was to be collected from students, ethical issues were also addressed and procedures developed to prevent those involved in the study being harmed as a result of their participation.

1.7 The Structure of the Thesis

Chapter 1 provides an introduction to the study including its aims, an introduction to the concept of metaknowledge, including its potential implications for professional practice and details of the context of the research, as well as issues prompting the author's initial interest in the issues addressed in the study. Chapter 2 provides a review of relevant literature, which is used as a basis for establishing specific research questions to be addressed later in the study. It initially provides theoretical context for the research by establishing the conceptual framework on which it is based. Alternative approaches for testing knowledge monitoring accuracy are also discussed in order to clarify subsequent discussion of findings from previous studies. To provide an insight into how findings which may arise from this study could be addressed in practice, potential reasons for those detected in previous research are considered, as well as remedial strategies which may be used to address their consequences. After establishing specific research questions and hypotheses on which the remainder of the study will focus, the chapter concludes by discussing how the study relates to previous research.

Chapter 3 addresses methodology and the research design to be used to answer the research questions established in the preceding chapter. It is important that the means used to derive findings and the assumptions under which they were produced are made clear (Jankowicz 1995). This assists readers in their interpretation of the research, by providing an indication of potential biases in how data have been collected and interpreted, as well as any other limitations in the work. The chapter commences by explaining the philosophy underpinning the research and the methodology adopted, before addressing the research design in detail. The latter includes an explanation of the nature of the research instrument and data collection procedures, as well as issues potentially affecting the reliability and validity of the research and initiatives implemented to address these. Data analysis procedures are also discussed and finally, ethical issues addressed.

Chapter 4 presents the findings arising from the application of the research design explained in Chapter 3 to the research questions established in Chapter 2. It commences by analysing the participants in the sample and the extent to which they are representative of the population from which they were drawn. Findings arising from analysis of the data are then discussed, as well as their potential implications and how they relate to those generated by previous studies discussed in Chapter 2.

In Chapter 5 the thesis concludes with a discussion of how the findings from the study contribute to the central theme of the research and importantly, their potential implications for professional practice. This includes discussion of the challenges they present for business schools, as well as specific initiatives which could be implemented. Since it is also important to recognise that findings are affected by limitations in the research process, these are also specifically addressed in this chapter, as well as the boundaries of the study and suggestions of possible directions for future research.

2. Literature Review

2.1 Introduction

Fixed research designs, such as that employed in this study, are theory driven and consequently, before embarking on the research, it is essential to have a good understanding of relevant theory and the phenomena under investigation, in order to determine the variables which are to be included in data collection procedures. This is accomplished by reviewing relevant literature to identify variables and relationships which may be investigated (Robson, 2002) and this chapter does so in respect of this study. Conceptual frameworks can be used to clarify and structure research by providing a 'map' of the field of study in question (Fisher, 2004 p.122) and in this case a framework is developed by initially defining concepts relating to metacognition before considering how these relate to each other. The potential implications of metacognition for learning are considered, as well as the specific relevance for this study of one aspect, metacognitive self-monitoring. Alternative approaches which may be used to test for metaknowledge and the manner in which it may be operationalised, are addressed to clarify subsequent discussion of previous studies addressing how accurately individuals are able to monitor their own knowledge. This includes research into individual differences as well as the association between metaknowledge and academic achievement. Since the potential implications of findings from this study for educational practice are to be addressed later, reasons for previous findings are also considered, as well as strategies which may be used to address their consequences. The chapter concludes by using the preceding discussion as a basis for establishing hypotheses to be tested and discussing how the study relates to previous research.

2.2 The Meta Concept

In discussing the development of a metacognitive model, Nelson (1996) explains that a conceptual analysis can produce theoretical paradoxes. For example, in respect of the concept of introspection, which is necessary for the self-monitoring of knowledge on which this study focuses, he cites the 19th century French philosopher Comte, who argued that one of the problems with introspection is that a thinker cannot divide himself in two parts, one of which reasons and the other observes, because the same organ cannot simultaneously take on the role of observer and observed. Nelson (1996) refers to an analogy used by Wilhelm Wundt, the founding father of experimental psychology, who likens this problem to a baron trying to remove himself from a bog by pulling on his own pigtail. He explains that psychologists addressing this issue have used a similar illustrative problem, in which two things appear to occur simultaneously, known as the liar's paradox. This relates to Epimenides, the ancient Greek philosopher who originated from Crete and is reputed to have stated that 'All Cretans are liars'. The problem here is that if this sentence is true then, as it was uttered by a Cretan, it must simultaneously be false. Nelson (1996 p.105) explains that the liar's paradox was not resolved until the middle of the 20th-century, when the philosopher Alfred Tarski developed the meta concept, in which meta refers to 'whatever about whatever' and the meta level is seen as being separate from the object level to which it refers. Thus, for example, if we consider analysis (the object level), the meta level is meta-analysis (i.e. analysis of analysis). Nelson (1996 p.105) uses the following sentence to illustrate how the meta concept can be used to resolve self referential paradoxes, such as the liar's paradox:

Thiss sentence contains threee errors

The problem here is that as the sentence contains only two spelling errors, the statement may at first glance appear to be false. However, the fact that it erroneously claims that there are three errors when there are two can be viewed as another error, which confirms the accuracy of the statement asserting that there are three errors, despite the fact that the sentence only appears to contain two. This difficulty can be resolved using Tarski's

theory of truth, by identifying two different statement levels. The first is the object level, at which sentences refer to subjects other than sentences and the second is the meta level, at which sentences relate to sentences. Thus, in the above sentence there are two errors at the object level (the misspelling of 'this' and 'three'), and one error at the meta level (the fact that the sentence contains two errors and not three). Nelson (1996) argues that a meta approach can also be applied to Comte's introspection paradox in that, rather than a single process, two simultaneous processes occur in introspective activity. The first occurs at an object level and relates to cognition about external objects. Cognition is defined by the Oxford English Dictionary as 'the action or faculty of knowing; knowledge, consciousness; acquaintance with a subject' and includes recalling and recognising knowledge, as well as intellectual abilities (Bloom, 1956). The second process in introspection occurs at the meta level and relates to cognition about those first level cognitions. This concept is known as metacognition.

2.3 Metacognition

The term 'metacognition' was first used by Flavell and Brown in the late 1970s (Paris and Winograd, 2003). While cognition is concerned with mental processes, such as thinking and learning, metacognition is 'knowledge and cognition about cognitive phenomena' (Flavell, 1979 p.906). Thus, it is a process through which cognitive processes are applied to themselves and has been described as one of the most intriguing issues in modern psychology (Yzerbyt, Lories and Dardenne, 1998). Schraw (1998) cites Garner's (1987) view that there is general agreement among researchers in the field that cognition and metacognition can be distinguished in that while performing a task requires cognitive skills, understanding how it was achieved requires metacognition. It has been likened to the idea that our minds contain a type of 'cognitive executive' that monitors thought and problem solving, to monitor its progress and consider how it can be assisted (Smith, Shields and Washburn, 2003 p.318). Learning is managed through a control process in which information is exchanged between the lower object level, at which cognitive activities occur, and the higher meta level, which manages these activities (Veenman, Van Hout-Wolters and Afflerbach, 2006) as illustrated in Figure 1. Thus, where an individual fails to understand something

at the object level, self-monitoring activity can inform control processes at the meta level, which prompts remedial learning strategies.

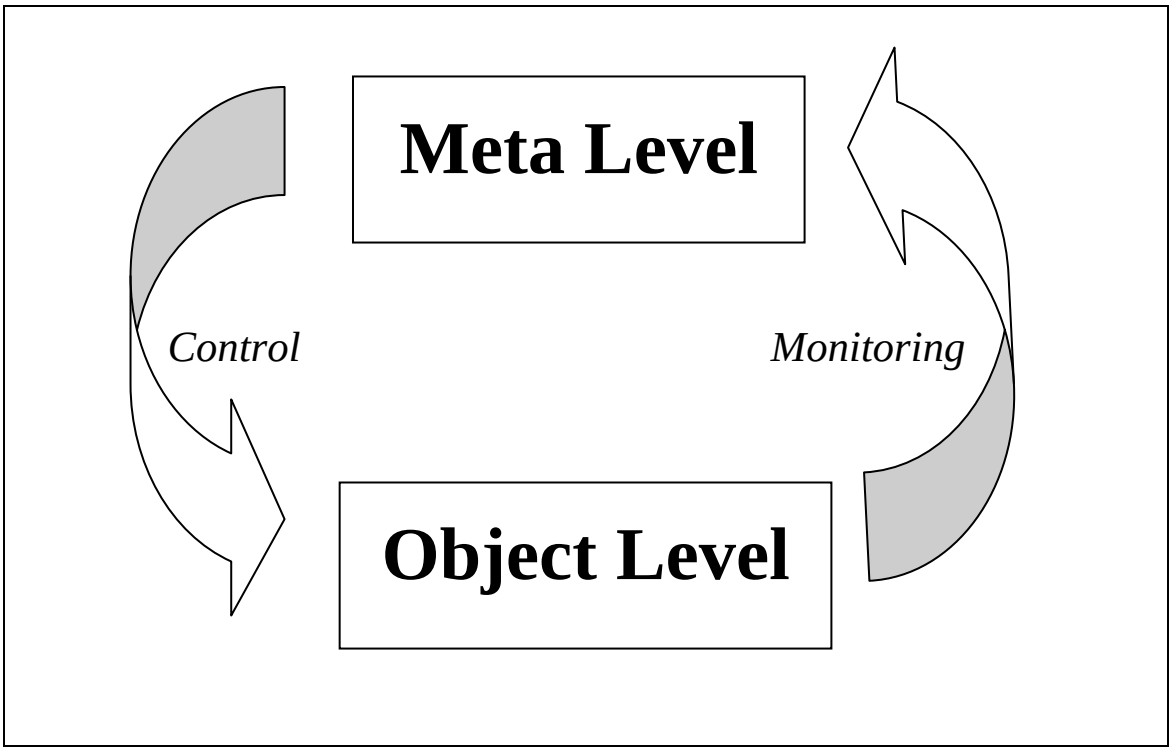


Figure 1: The Metacognitive Control Cycle

Source: Adapted from Nelson (1996)

Metacognition is therefore concerned with the appraisal and management of learning, through monitoring, evaluation and planning (Everson and Tobias, 1998). It has been described as one of the key differences between human and animal cognition, with the very existence of psychology, providing evidence of our interest in mental procedures (Lories, Dardenne and Yzerbyt, 1998). However, the view that it is exclusively a human capacity is now being challenged, with recent work attributing metacognitive ability to more intelligent animal species, such as monkeys and dolphins (Smith, Shields and Washburn, 2003).

While there is general agreement about the importance of metacognition, Veenman, Van Hout-Wolters and Afflerbach (2006) explain that its conceptualisation lacks coherence and argue that more work should be done to define it and its components more consistently. For example, the concept of self-regulation, which is commonly associated with metacognition, is considered by some to be a component of metacognition, while others consider that it is super-ordinate to it. Many researchers do not use formal definitions but, rather, prefer to use examples to illustrate metacognition and Paris and Winograd (1990) attribute this to the fact that as *any* cognition of thinking could be classed as metacognition, definitions would have to be relatively open-ended. They explain that this can give rise to disagreement about its nature with, for example, some regarding it as conscious activity and others arguing that it can be unconscious and inaccessible. Hacker (1998 p.3) explains that, while it may be considered a 'fuzzy concept', the notion of individuals thinking about their thoughts is fundamental. This thinking may relate to what they know ('metacognitive knowledge'), what they are doing ('metacognitive skill') or their current cognitive state ('metacognitive experience') and these aspects have been included in various theoretical accounts of metacognition. He defines metacognition as 'thinking about thinking' or 'cognition of cognition' and distinguishes metacognitive thought from other types of thinking, based on its source, in that it does not arise from an individual's 'immediate external reality' but rather from their 'own internal mental representation of that reality'.

2.3.1 Flavell's Conceptual Model

As explained earlier, Flavell was one of the first to use the term metacognition, in the late 1970s. Hacker (1998 p.4) explains that his early work was influenced by the developmental psychology of Jean Piaget, who distinguished between first and second degree operations. The former, which relate to 'thoughts about an external empirical reality' may become the subject of the latter, which are therefore of a higher order. It is this higher level activity with which metacognition is concerned. Flavell (1979) suggested that it relates to the interaction between four phenomena:

1) Goals and tasks
2) Actions and strategies
3) Metacognitive knowledge
4) Metacognitive experiences

Goals and tasks relate to the objectives of cognitive activity, whereas actions and strategies are the means through which these are achieved and are influenced by metacognitive knowledge and experiences.

2.3.1.1 Metacognitive Knowledge

Flavell (1979) divides metacognitive knowledge into further subcomponents, which he describes as being related to person, task and strategy. The 'person' category embraces understanding of how one (and others) learns and can be broken down further into knowledge about intra-individual and inter-individual differences, as well as universal beliefs. Intra-individual differences could include, for example, an individual's belief that s/he learns better with active rather than passive experiences, whereas inter-individual differences arise from a belief that individuals may react differently in response to a particular stimulus. Universal beliefs could include an appreciation that failure to fully commit to learning generally tends to hamper it. The 'task' category entails an understanding of the requirements of different tasks and an appreciation that some are likely to be more demanding than others, with for example, some being better informed by supportive information which may be available. This suggests that person and task knowledge may interact to inform learning in that appreciation of task difficulty for example, may be informed by the extent to which an individual's preferred learning style is likely to be effective in tackling it. The 'strategy' category is concerned with an individual's ability to discriminate between alternate approaches for achieving a particular goal and choose the most appropriate. The potential for combining this with other metacognitive knowledge is again apparent in that for example, the most appropriate learning strategy for an individual will be informed by not only their

appreciation of the alternative possibilities that exist, but also which ones are likely to be most effective for them.

2.3.1.2 Metacognitive Experiences

Flavell (1979) suggests that individuals have 'metacognitive experiences' in which metacognitive knowledge enters the consciousness, before, during, or after an intellectual activity. These experiences are likely to be more common in cognitive activity which requires much careful thought, such as new learning situations, and play an important role in the ongoing development of metacognitive knowledge, in that individuals are able to modify and add to it, in response to them. Flavell (1979) explains how the four phenomena outlined above interact so that for example, when assigned a task with a particular goal, an individual may draw on current metacognitive knowledge, resulting in a metacognitive experience indicating that the task is likely to be difficult. Further metacognitive knowledge can then be applied which may, for example, result in obtaining further information about the task before commencing it. Having commenced it, a subsequent metacognitive experience may indicate how it is proceeding and this, combined with further metacognitive knowledge, may prompt modifying the approach to the task, in order to achieve the ultimate objective more easily.

2.3.2 The Knowledge and Regulation Conceptual Framework

While Flavell's model provided an early representation of metacogniton and its components, the most common distinction in the literature is that between metacognitive knowledge and skills (Veenman, Van Hout-Wolters and Afflerbach, 2006). Paris and Winograd (1990 p.16) refer to a distinction between 'knowledge about cognitive states and processes' and 'cognitive self-management'. These have also been described as knowledge about cognition and regulation of cognition (Everson and Tobias, 1998), and relate to understanding and controlling cognitive processes respectively. Schraw (1998) offers a framework for understanding metacognition which, while differing from Flavell's (1979) earlier representation to some extent, includes common and similar features. It is based on the commonly used distinction between

knowledge about cognitive states, and the control, or self management, of cognitive enterprises, as shown in Figure 2.

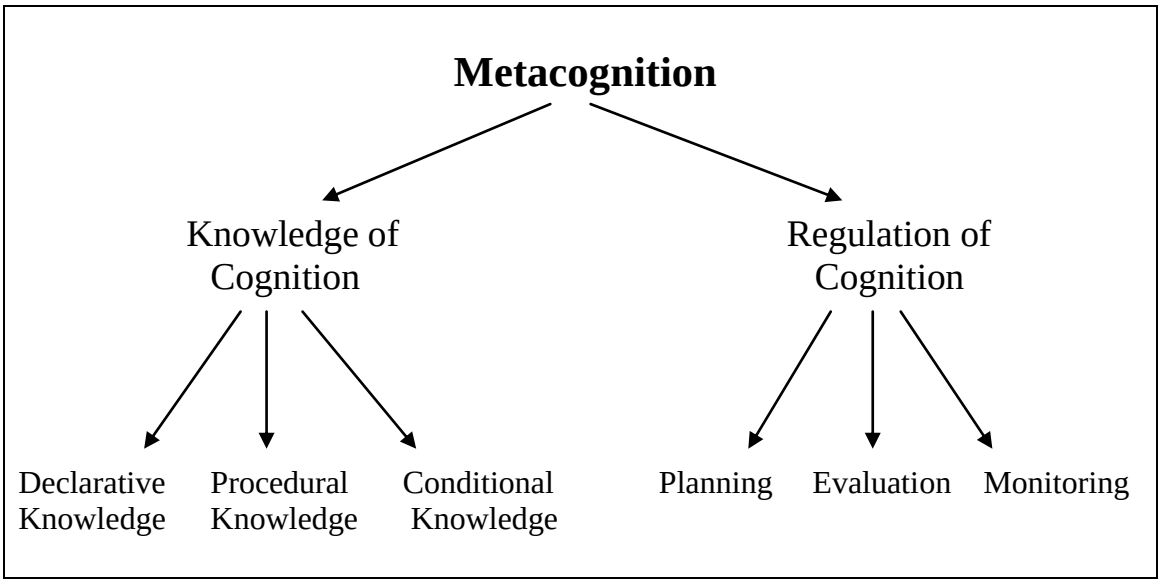


Figure 2: A Conceptual Framework for Metacognition
Source: Based on Schraw (1998)

2.3.2.1 Knowledge of Cognition

Knowledge of cognition relates to what one knows about one's own cognition, or about cognition generally. This relates to the metacognitive knowledge to which Flavell referred and can be used to 'select, evaluate, revise and abandon cognitive tasks' (Flavell, 1979 p.908). It can be divided into three elements: declarative, procedural and conditional knowledge.

Declarative Knowledge

Hacker (1998 p.8) refers to a definition of declarative knowledge by Kluwe (1982 p.203), as 'stored data in long-term memory'. In the context of metacognition, it relates to knowledge about approaches which may be used for cognitive tasks, such as the use of skim reading to develop appreciation and the use of diagrams to enhance

understanding (Schraw, 1998). It is also concerned with an individual's understanding of how they learn and the factors that impact on their success and in this respect, relates closely to Flavell's (1979) 'person' category. Pintrich (2002 p.221) refers to this aspect as 'self knowledge' and includes within this an appreciation of one's strengths and weaknesses, an important aspect of which is 'self-awareness of the breadth and depth of one's knowledge'. He claims that a surprisingly high number of college students lack accurate self-knowledge and that this can constrain their learning. As well as including an individual's appreciation of both how they think and what they know, Paris and Winograd (1990) also advocate broadening the scope of this aspect of metacognition to include motivational aspects of learning. Pintrich (2002) supports this view, arguing that self knowledge should incorporate an individual's self awareness about their motivation, which relates to how interested in a task they are and how much they value it, and suggesting that there are important links between these issues and learning.

Procedural Knowledge

Procedural knowledge is concerned with how to use the various approaches of which one may have declarative knowledge to perform different tasks and equips individuals with a range of methods which may be used in different contexts. Examples include an appreciation of how setting goals can assist in planning learning, understanding that asking oneself whether a mathematical calculation has been performed correctly can help to monitor learning and appreciating that, if necessary, taking steps to amend the calculation may usefully regulate learning (Schraw, 1998). Such appreciation has also been referred to as 'strategic knowledge', which relates to generic strategies applicable across different tasks and domains (Pintrich, 2002 p.220). Thus, the term 'strategic' is applied inconsistently, with Flavell (1979) using it to refer more explicitly to the ability to discriminate between different approaches, something which Schraw (1998) categorises separately from procedural knowledge and refers to as conditional knowledge.

Conditional Knowledge

Conditional knowledge is concerned with appreciating when particular aspects of declarative and procedural knowledge should be used in determining the appropriate response to a particular task (Schraw 1998). It enables an individual to choose from a number of strategies available to tackle the cognitive task and therefore, as explained above, relates closely to Flavell's (1979) strategy category. It can be particularly powerful, as it helps to ensure that existing knowledge is used appropriately and effectively. Conditional knowledge is considered by some to be declarative knowledge and by others, to be an aspect of metacognitive skill rather than knowledge, which provides further evidence of the lack of coherence in the conceptualisation of metacognition and inconsistency in the definition of its components (Veenman, Van Hout-Wolters and Afflerbach, 2006).

The three elements of knowledge of cognition to which Schraw (1998) refers can therefore be summarised as referring to awareness of various strategies (declarative knowledge), how they can be used (procedural knowledge) and when they should be used (conditional knowledge). As discussed above, they may be used interactively and Flavell (1979 p.907) argued that most metacognitive knowledge is concerned with combining them. He illustrates this with a scenario in which an individual believes that:

they	*(unlike one of their peers),*	-	*Declarative knowledge*
should use strategy A	*(rather than strategy B),*	-	*Procedural knowledge*
in task X	*(but not in task Y)*	-	*Conditional knowledge*

Schraw (1998) advocates the use of strategic evaluation matrices (SEM) to assist learners in developing their knowledge of cognition (See example in Appendix 1). These can help them to develop their declarative, procedural and conditional

understanding of various strategies, which helps them to appreciate how and when they may be used, and for what purpose. This understanding can then be applied in regulatory activities to enhance learning.

2.3.2.2 Regulation of Cognition

Regulation of cognition, which has also been referred to as 'cognitive self-management', relates to activities used by individuals to control their learning and includes planning and implementing appropriate strategies for learning, as well as monitoring and adapting learning as it proceeds (Paris and Winograd, 1990 p.18). Sternberg (1998) argues that metacognitive processes for learning and using information are as important as knowledge itself and while other skills have been identified in the literature, the most commonly discussed are planning, evaluation and monitoring (Schraw, 1998).

Planning

Planning entails scheduling sufficient time for various learning tasks, in order to attain learning outcomes. This regulatory activity can be informed by knowledge of cognition to promote effective learning. For example, as illustrated in Figure 3, learners can plan to carry out learning using strategies of which they are aware and have procedural knowledge and which their conditional knowledge indicates are appropriate for the task in hand.

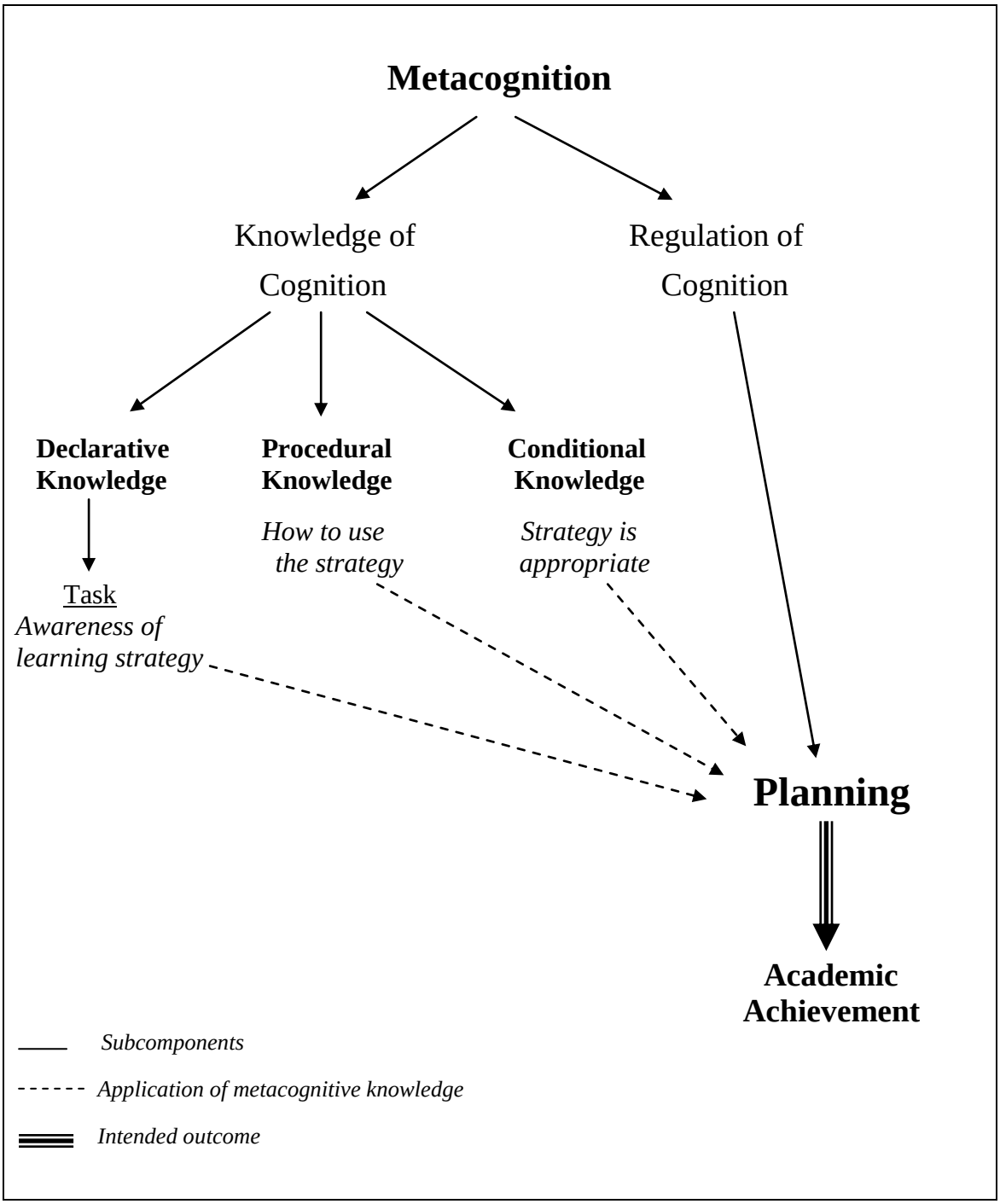


Figure 3: The Role of Metacognitive Knowledge in Planning Learning

Evaluation

Evaluation refers to the self-assessment of learning outcomes and includes critically assessing learning outputs to judge their value. Klenowski 1995 (p. 146) defines such

self-assessment as 'the evaluation or judgment of "the worth" of one's performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes'. Falchikov and Boud (1989) stress the educational benefit of this type of activity, the value of which lies in its propensity to encourage a greater degree of reflection on learning (Mowl & Pain, 1995) and Ward, Gruppen and Regehr (2002, p.76) argue that its importance in education 'stands undisputed'. A common approach in studying this aspect of metacognitive self-management is to investigate learners' ability to provide an accurate self-assessment of their academic output, by forecasting their assessment grades (Boud and Falchikov, 1995).

Monitoring

Monitoring has been defined as 'one's online awareness of comprehension and task performance', of which self-testing during learning is an example (Schraw, 1998, p.115). This definition suggests that it can be distinguished from evaluation, in that it refers to awareness of how one is performing while carrying out a task (i.e. online) as opposed to self-evaluation of the outputs or products of learning. However, as discussed earlier, components of metacognition tend to be defined inconsistently. Paris and Winograd (1990) for example, classify reflection on one's knowledge state, which could be included in Schraw's (1998) definition of monitoring, as cognitive self-appraisal, which they distinguish from self-management of cognition, in which they include monitoring ongoing performance. Nelson (1996) meanwhile, when identifying different strands of monitoring research, includes prospective activity, such as predicting how easy future learning will be, and this does not comply with the online aspect of Schraw's (1998) definition. Consequently, while metacognition clearly embraces the notion of self-assessment of both ongoing learning and the manner in which it is used to produce outputs, there are differences in how these are classified. As Veenman, Van Hout-Wolters and Afflerbach (2006) argue, there is clearly a need for more precise taxonomies of metacognitive knowledge and skills.

In terms of how regulation of cognition can enhance learning, research indicates that better appreciation of gaps in understanding is important (Schraw, 1998). Learning requires that, as well as working independently, students are able to assess their own progress (Falchikov & Boud, 1989) and therefore accurate self-monitoring is essential

for effective learning (Pieschl, 2009). Consequently, monitoring of knowledge is considered to be a particularly important component of metacognition, as without accurate monitoring, learning cannot be effectively controlled, particularly in environments which require large amounts of information to be learned (Clarebout, Elen & Onghena, 2006). This implies that it may be particularly valuable in a higher education learning environment.

To illustrate how self monitoring may impact on behaviour, Dunning et al (2003) report a study in which students answered a series of questions based on scientific reasoning. While females provided less favourable self-assessment of their performance, there was no significant difference in actual test performance. However on being invited to participate in a further quiz, for which prizes were offered, females were less inclined to accept than males and Dunning et al (2003) therefore suggested that it is perception rather than reality which influences behaviour. This conclusion can be challenged on the grounds that, rather than perception of performance, other gender-related factors, which were not addressed in the study, may have influenced the behaviour of the participants. For example, their greater reluctance to participate further may have been attributable to the prizes on offer for doing so being less appealing to the female participants. However, if perception does influence behaviour as suggested, then misconceptions may produce inappropriate responses and poor self-monitoring may, for example, result in those who over-estimate their current knowledge state, failing to address this unappreciated deficiency in future learning activities. Flavell (1979) argues that those with better metacognitive monitoring skills are likely to learn more effectively, both inside and outside the classroom environment, and suggests that they may be better equipped when taking lifestyle decisions, as well as learning more effectively in formal educational situations.

In terms of the relationship between components of metacognition, as illustrated in Figure 4, declarative self knowledge relating to appreciation of one's own knowledge can be used to inform monitoring activity, through which gaps in knowledge may be identified. This may in turn prompt regulatory planning activity, which can also draw on other metacognitive knowledge, and initiate learning strategies aimed at addressing the knowledge deficit and enhancing academic achievement.

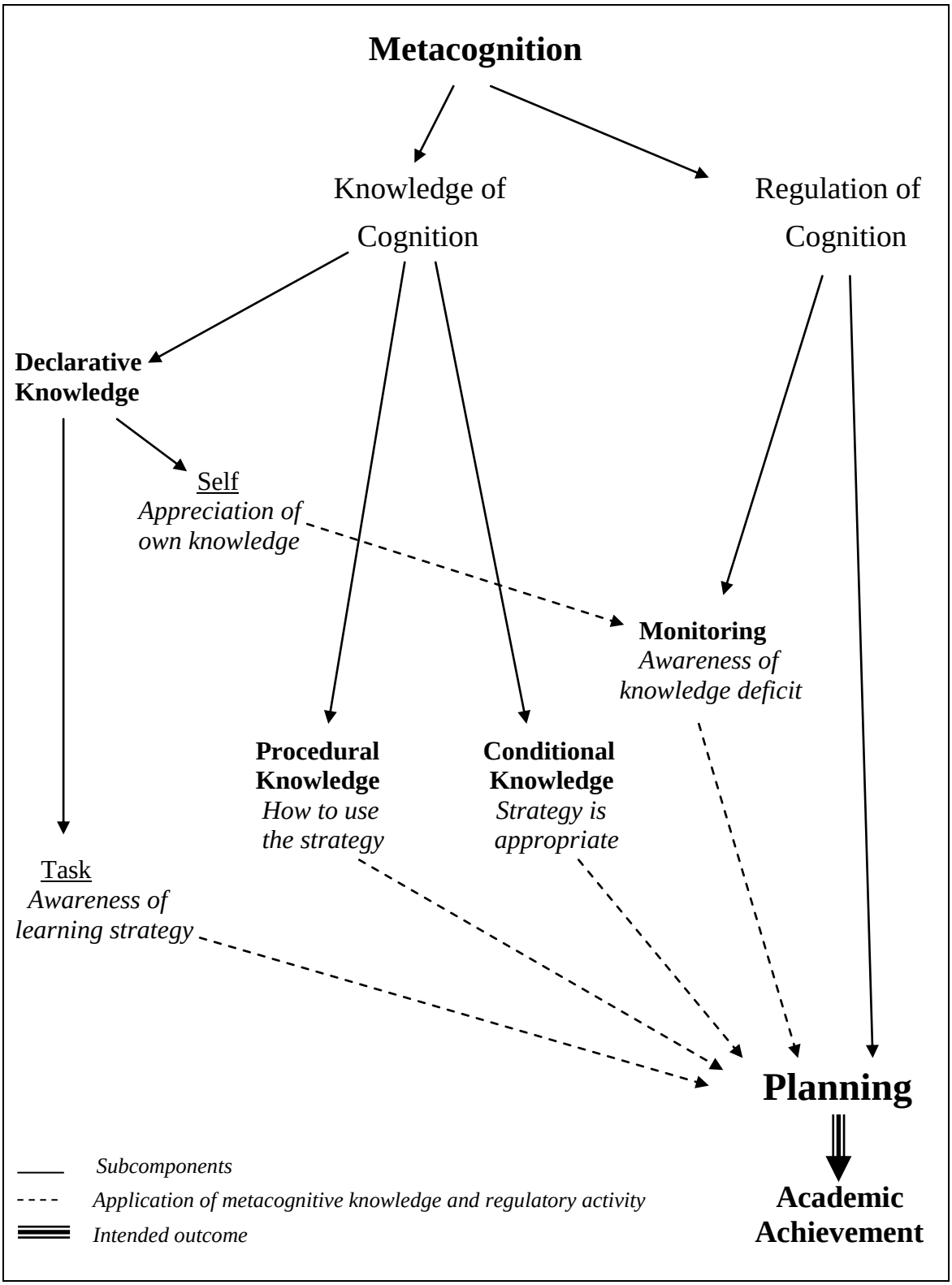


Figure 4: Interaction between Monitoring, Metacognitive Knowledge and Planning

However, Nelson (1996, p.106) explains that, while philosophers have traditionally assumed infallibility in introspection, metacognitive theory does not and assumes that 'cognitive illusions' occur, through individuals' imperfect monitoring of their own cognitive activity. He describes how these can be investigated by eliciting meta-level statements about object level cognitive activities, operationalising what actually occurs at the objective level and testing for self-monitoring accuracy, by assessing the relationship between the two. This approach has been commonly adopted in research studies using responses to a series of multiple choice knowledge questions at the object level, accompanied by respondents' meta-level judgements as to their perception of the accuracy of their responses as illustrated in Figure 5.

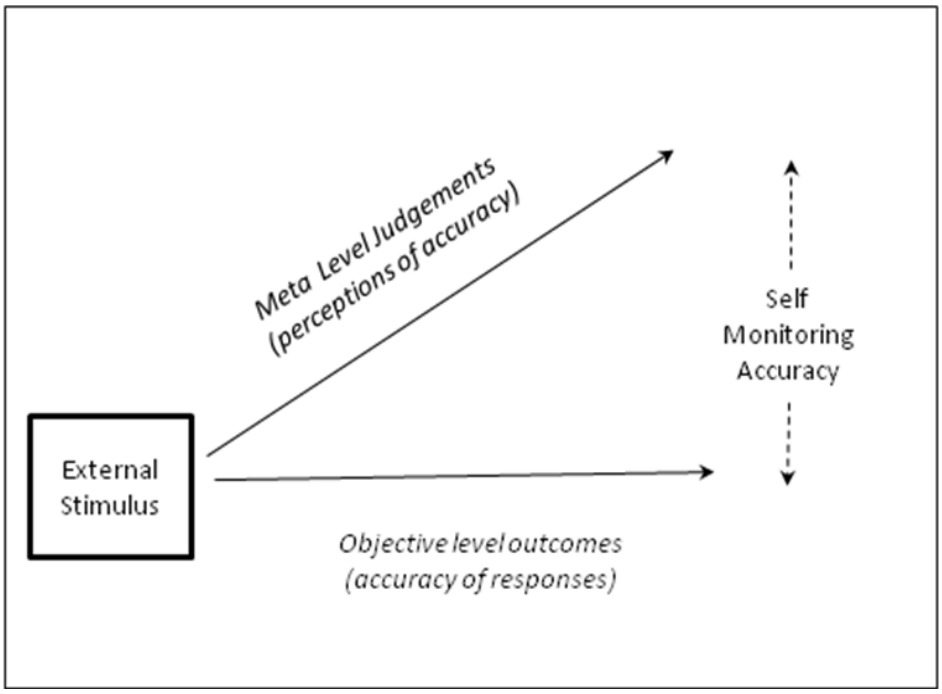


Figure 5: Assessing Self-Monitoring Accuracy

Source: Adapted from Nelson (1996)

Schraw (1998) argues that, while the strategy evaluation matrices discussed earlier may improve knowledge of cognition, they may not necessarily improve regulatory skills. To do so, he advocates the use of a regulatory checklist (see example in Appendix 2), which provides useful prompts to encourage students to be more strategic and systematic and has proved to be effective in learning.

2.3.3 Metacognition and Learning

The raised consciousness of thinking associated with metacognition promotes self regulated learning as it transfers responsibility for monitoring of learning from teacher to learner and also promotes positive self-perception in the learner while increasing motivation to learn (Paris and Winograd, 1990). Thus, learning becomes more independent as learners participate more actively in their own learning, rather than over-relying on pre-determined activities, in which they are instructed and tend to play a more passive role. Since it emphasises personal knowledge and self-management, it also embraces individual differences in learning styles and cognition and has therefore been linked to successful development of expertise through learning (Sternberg, 1998). While it has been found to be only moderately correlated with intelligence, there is much evidence indicating that metacognition complements intellectual ability in enhancing learning and that metacognitive skill can therefore compensate for poor cognitive ability (Veenman, Van Hout-Wolters and Afflerbach, 2006). Consequently it has the potential to enhance academic performance (Schraw, 1998) and Everson and Tobias (1998) explain that studies have shown that those with better metacognitive skills are generally better equipped to monitor their learning accurately, assess the extent of their knowledge, plan for new learning and enhance their knowledge. However, they argue that, while this is particularly important for effective learning in dynamic environments, in which available information is continually updated, many learners have ineffective metacognitive strategies. Consequently, an important message arising from research in the early 1990s on competent thinking was that students should be made aware of different strategies and given the chance to develop metacognition about them (Pressley and Gaskins, 2006). Paris and Winograd (1990) advocate creating opportunities in the curriculum for students and tutors to discuss beliefs and judgements in respect of academic activities, since this will enable learners to reflect on their current approaches to learning and how metacognition may assist them in enhancing them.

2.3.3.1 Metacognitive Knowledge and Bloom's Taxonomy for Learning

In 1949 Benjamin Bloom initiated the compilation of a taxonomy of educational objectives, primarily as a means of providing a consistent framework for the production and exchange of test questions between American universities (Krathwohl, 2002). This resulted in collaboration over the next few years which culminated in the publication of the taxonomy and an accompanying handbook in 1956 (Bloom, 1956 – see Appendix 3). This gained widespread acceptance and was translated into 22 languages and while initially primarily conceived as a tool to assist in measuring educational attainment, Bloom believed it could also be used to help clarify the meaning of educational goals (Krathwohl 2002). Since its inception, it has been used to help ensure that learning objectives include more complex elements of the taxonomy, such as synthesis and evaluation, as well as the acquisition of knowledge.

However Klenowski (1995), in advocating initiatives to help develop learners' self-evaluation skills, argued that they should be encouraged to develop the ability to make judgements earlier than Bloom's taxonomy suggested and a revised version was later produced (Anderson et al, 2001). It was devised by a team comprising cognitive psychologists, instructional researchers and assessment specialists, including one of the authors of the original taxonomy, David Krathwohl. The revised version included numerous refinements, such as changing the category names from nouns to verbs, to reflect the manner in which learning objectives are more typically framed. Thus, for example, application became apply and evaluation became evaluate. However, one of the most significant changes was to convert the original one dimensional framework to a two dimensional model which acknowledged the importance of metacognitive knowledge. This two dimensional approach arose from the typical framing of learning outcomes in terms of both subject content and a cognitive process, relating to what the learner is expected to do with that subject matter. Hence, they usually entail the use of a noun, or noun phrase, representing the subject matter and a verb, representing the cognitive process. For example, a research methods course may include the following learning outcome:

The learner will be able to apply….	*(verb – cognitive process)*
data collection methods	*(noun phrase – subject content)*

Krathwohl (2002) explains that, in the original taxonomy, the knowledge category differed from the others in that it had two dimensions, as it included both noun and verb elements. The verb aspect was included in the definition of knowledge in the handbook accompanying the taxonomy, which referred to behaviours emphasising '*recognition* or *recall*, of ideas, materials or phenomena' (Bloom, 1956 p.62, italics added for emphasis), while the noun aspect was incorporated in the various sub-categories of knowledge to which these behaviours could be applied (terminology, facts, conventions etc.). Krathwohl's (2002) suggestion that this two-dimensional aspect related only to the knowledge category, can be challenged in that this was in fact also the case for the analysis category which referred to analysis (verb aspect) of elements, relationships and principles (nouns).

This two-dimensional approach was specifically incorporated in the revision to the taxonomy, in which the verb aspect forms one dimension, based on cognitive processes, and the noun aspect another, based on knowledge (see Appendix 4). The cognitive process dimension comprises six progressive, but overlapping, major categories (remember, understand, apply, analyse, evaluate and create), which are broken down into a further 19 sub-categories as shown in Appendix 5. The knowledge dimension was structured with four categories, as compared to the three used in the original taxonomy (see Appendix 6). The first three, – factual knowledge, conceptual knowledge and procedural knowledge – consolidated sub-categories used in the original framework. However, the fourth – metacognitive knowledge – added an aspect which was not recognized at that time, but which has become increasingly significant due to research which has demonstrated how important it is to make learners aware of how knowledge of their metacognitive activities can be used to adapt their thinking and learning effectively (Krathwohl, 2002). It includes three sub-types – strategic knowledge, knowledge about tasks and self-knowledge – which relate to Flavell's (1979) strategy, task and person sub categories of metacognitive knowledge. The importance of metaknowledge is specifically acknowledged in the revised taxonomy in illustrative examples of self-knowledge provided, which include 'awareness of one's own knowledge level' and the need for educators to assist students in making accurate assessments of their knowledge is also emphasised (Anderson et al, 2001 p.29).

2.3.4 Metacognitive Monitoring Research

Hacker (1998) explains that three strands of research consistently appear in metacognition research. The first is concerned with cognitive monitoring and studies in this field have investigated how accurately individuals are able to monitor their knowledge and thought processes. The second is cognitive regulation, where studies have typically addressed the ability to transfer learning strategies between different tasks. Note here further evidence of inconsistency in the definition of various aspects of metacognition in that, while Schraw (1998) classifies monitoring as a regulatory process, Hacker (1998) distinguishes between regulation and monitoring. The third research strand addresses both monitoring and regulation and includes studies which investigate people's ability to monitor their own thinking and use this information in subsequent activities (Hacker, 1998). This study focuses on the first of these, monitoring, and Figure 6 shows types of research which have been carried out in this strand, during the acquisition and retrieval of knowledge. A distinction can be made here between prospective activities, which entail predicting future performance, and retrospective activities, which entail postdiction (Nelson, 1996).

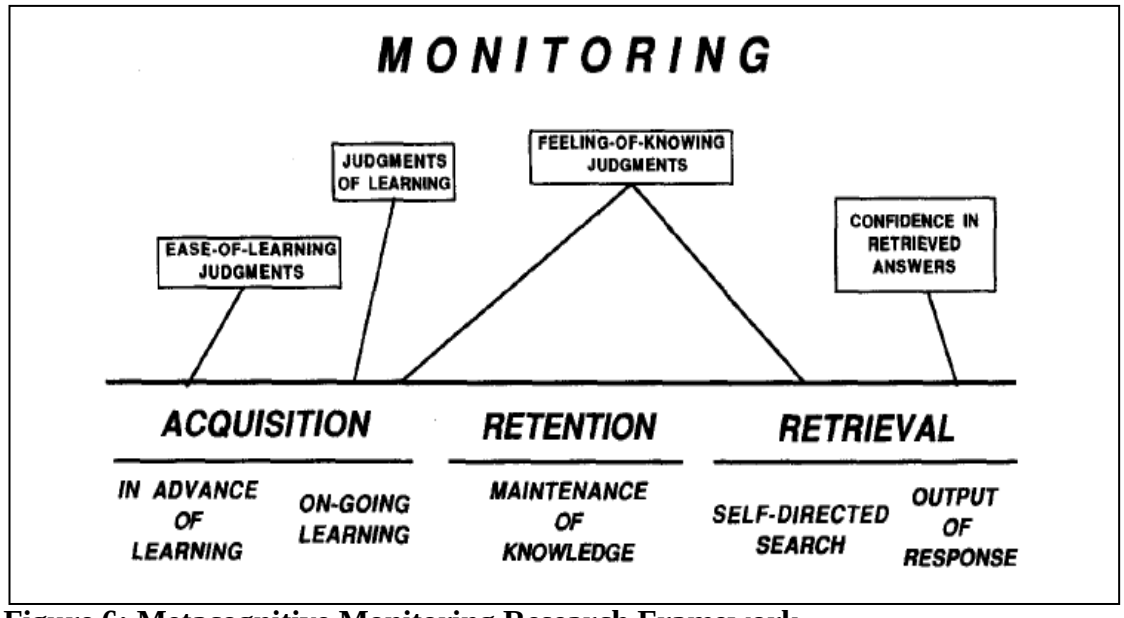


Figure 6: Metacognitive Monitoring Research Framework
Source: Nelson (1996 p.109)

2.3.4.1 Prospective Activity

Prospective activities at the acquisition level include predicting how easy learning is expected to be (ease of learning judgements) and how easily the learner expects to be able to subsequently recall ongoing learning (judgements of learning). Straddling both the acquisition and retrieval stages are 'feeling of knowing' judgements, which relate to an individual's ability to subsequently recognise knowledge items which they cannot currently recall. Research in this aspect has investigated the extent to which individuals, who are currently unable to recall an item of knowledge, are subsequently able to do so on being presented with alternative possible correct answers.

2.3.4.2 Retrospective Activity

Retrospective activity relates to the post-learning retrieval stage and studies here have investigated the accuracy of meta-level judgements, in which individuals express levels of confidence in their current knowledge. Thus, metaknowledge can be assessed by investigating how accurately individuals are able to monitor their own knowledge. The importance of this aspect is emphasised by Hacker (1998 p.11), who argues that while there may be a lack of clarity in classifying some aspects of metacognition, there appears to be consensus that any definition should incorporate the concept of 'knowledge of one's knowledge'.

2.4 Metaknowledge

Russo and Schoemaker (1992 p.8) explain that while knowledge relates to 'the facts, concepts, relationships, theories and so on that we have accumulated over time', metaknowledge is 'an appreciation of what we do know and what we do not know'. Thus, those with well developed metaknowledge have a good appreciation of what they know and the gaps in their knowledge, whereas those with poor metaknowledge are less able to appreciate the state of their knowledge. This capability has also been referred to as 'knowledge monitoring ability' (KMA) (Everson and Tobias, 1998 p.66) and it is important to distinguish it from metacognitive knowledge which, as discussed above,

relates to ‘knowledge about cognition in general’ (Pintrich, 2002 p.219), a wider perspective than the ‘knowledge about knowledge’ to which metaknowledge relates.

The former US Secretary for Defence, Donald Rumsfeld (2002), referred to the capacity for appreciating gaps in knowledge, in his explanation that:

> *As we know, there are known knowns. There are things we know we know. We also know there are known unknowns. That is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know.*

He was much derided for this observation and indeed won The British Plain English Campaign’s annual prize for the most nonsensical remark made by a public figure, narrowly beating a quote from European Commissioner Chris Patten, who in 2003, claimed that ‘the British Conservative Party had committed political suicide and was now living to regret it’ (BBC, 2003). However, Rumsfeld’s quote contains a perfectly logical statement and our capacity to learn is dependent on the ability to reflect on, and develop awareness of, what we know and what we do not (McGregor, 2004). He was perhaps displaying the wisdom referred to in the following Arabic proverb cited by Davidoff (1995):

> *He who knows and knows that he knows is conceited; avoid him.*
> *He who knows not and knows not that he knows not is a fool; instruct him.*
> *He who knows and knows not that he knows is asleep; awaken him.*
> *But he who knows not and knows that he knows not is a wise man; follow him.*

Similar views have also been attributed to Confucius who reportedly explained that, ‘real knowledge is to know the extent of one’s ignorance’ (Dunning et al, 2003 p.83) and Thomas Jefferson, who suggested that, ‘he who knows best, best knows how little he knows’ (Kruger and Dunning, 1999 p.1130). Metaknowledge has been described as

‘higher-order knowledge’ (Sternberg, 1998 p.129), which is associated with a ‘higher level of expertise: understanding the nature, scope and limits of our basic, or primary, knowledge’ Russo and Schoemaker (1992 p.8). Gredler (2004) argues that it is essential when collecting data for solving problems and consequently, if well developed, it can assist in making more informed choices and decisions. However, since it requires a greater level of expertise than primary knowledge (Ramnarayan, Strohschneider and Schaub, 1997) and because it tends not to be formally recognised or rewarded, nor developed during formal education, metaknowledge can be poorly developed and this may remain a hidden problem (Russo and Schoemaker, 1992).

2.5 Conceptual Framework

Having considered metacognition and its various sub-components, as well as how they may relate to each other, the conceptual framework which will be used to guide this study is shown in Figure 7. This illustrates how metaknowledge, a form of declarative self-knowledge, informs regulatory monitoring activity, which can in turn prompt the planning of learning strategies to enhance academic achievement. Given its role in prompting planning of learning, accuracy of knowledge monitoring is an important factor and various methods may be used to test for this.

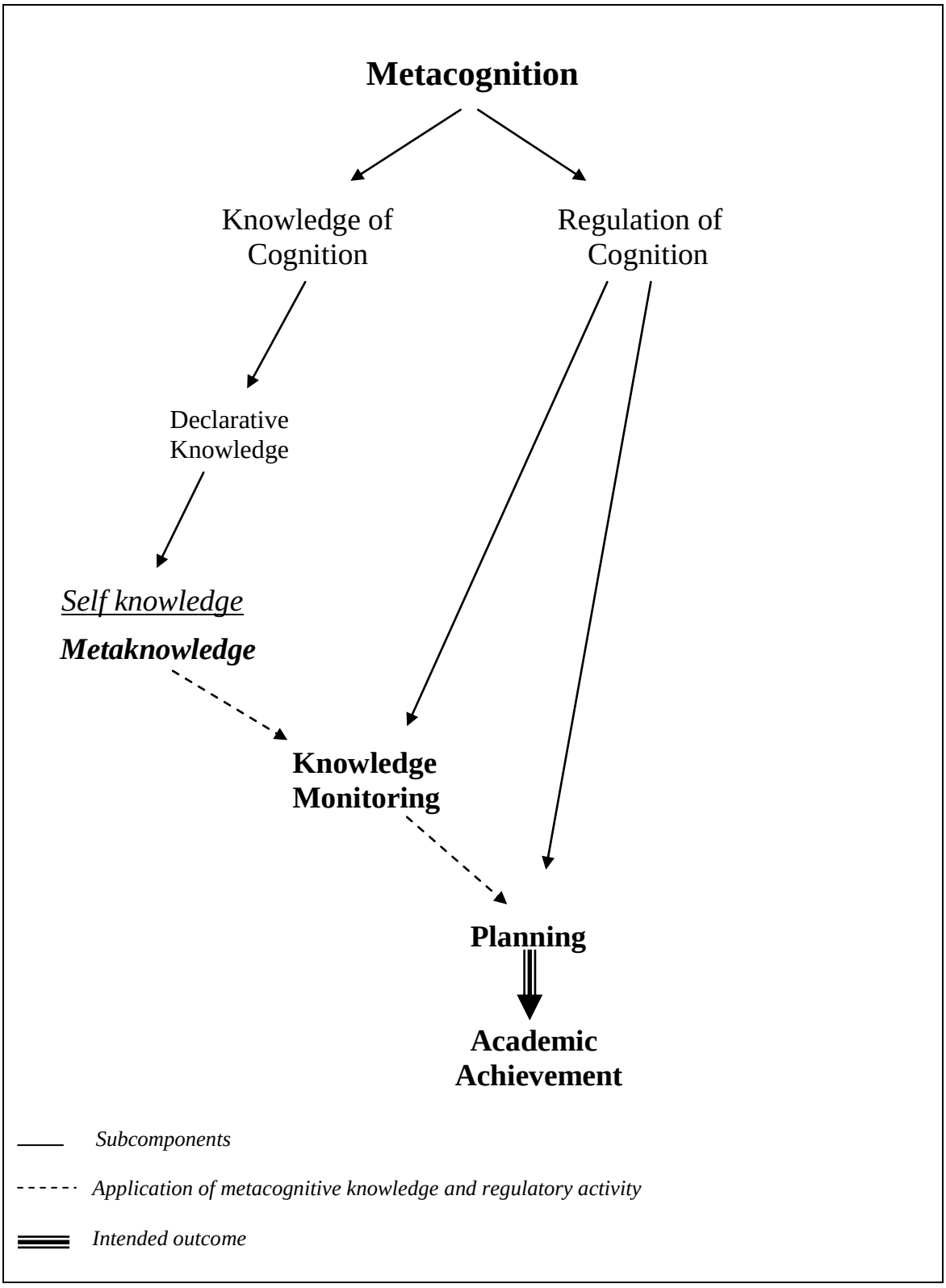


Figure 7: The Conceptual Framework Developed for This Study

2.6 Testing for Knowledge Monitoring Accuracy

Investigating the extent to which individuals are able to assess their own knowledge, requires asking them to make judgements about the accuracy of responses they provide to questioning designed to investigate it. To facilitate unbiased interpretations of these responses, objective questions, for which there is a definitive correct answer, can be used (Kruger and Dunning, 1999). When doing so, a distinction can be made between approaches in which probability assessments are made over continuous variables, such as unknown numerical values, and calibration studies that seek judgements on particular propositions Keren (1991).

2.6.1 Continuous Variables (Range Questions)

When using confidence range estimates, participants are provided with a series of questions, each of which requires a numerical response. They are asked to provide a lower and upper estimate for each question, such that they believe the correct answer has a given percentage chance of falling within this range (see example in Appendix 7 using 90% confidence ranges). The resulting data is analysed to reveal a 'surprise index', which reflects the percentage of judgements for which the correct answer lies outside of this range. These are compared with the anticipated error rate to provide a metaknowledge rating as follows:

Metaknowledge score = S.I. - T.E.R.

Where:

S.I. = *the surprise index (i.e. the rate of incorrect responses)*

T.E.R. = *the targeted error rate (i.e. in this case 10 %*)*

** In the example in Appendix 7, respondents were asked to make judgements at a 90% confidence level.*

A score of zero indicates a well calibrated respondent who, having being asked to provide answers with a 90% confidence level, did actually experience a 90% success rate. Thus, the surprise index in this case would coincide with the targeted error rate as follows:

Metaknowledge score	=	*S.I.*	-	*T.E.R.*
	=	*(100% - 90%)*	-	*10%*
	=	*10%*	-	*10%*
	=	*0%*		

A score of greater than zero indicates that the respondent was less capable of identifying the correct range than they believed and was therefore overconfident about how much they knew about the subject matter in question. To better reflect their lack of knowledge, the ranges they provided should have been increased. Conversely, a metaknowledge score of less than zero indicates underconfidence. In such cases the respondent's knowledge was actually better than they believed and consequently, the ranges provided could have been narrower to reflect this.

2.7 Calibration Studies

When using judgements on propositions, Schraw (2009) distinguishes between two approaches. In the first, respondents make a dichotomous judgement on an outcome, indicating whether they believe they will be successful or not. Their accuracy can be subsequently verified by comparing their actual and estimated success rates. In the second, more common, approach respondents express their confidence in the accuracy of their judgements using an appropriate continuous rating scale. This type of investigation is known as a calibration study. Calibration is a measure of the relationship between individuals' metacognitive judgements and their own performance and has been referred to in the literature by a number of terms including accuracy,

illusion of knowing and judgement bias (Pieschl, 2009). The most popular approach in calibration studies is one in which respondents are required to select from a number of *n* alternatives and give the subjective probability that their choice is correct (Keren, 1991). Using this method, a respondent without the knowledge to enable them to discriminate between the correct and incorrect alternatives available to them, may still do so by chance and therefore the confidence rating scale should be designed accordingly. If for example, there are two alternatives, a scale of 50-100% confidence should be used, since the respondent should not be less than 50% confident on chance grounds alone. Respondents are said to be well calibrated where 'for all propositions assigned a given probability, the proportion that is true is equal to the probability assigned' (Lichtenstein, Fischhoff and Phillips, 1982 p.307) as estimations which closely resemble the actual outcomes associated with those judgements demonstrate a good appreciation of the extent of their knowledge.

A visual representation of calibration can be provided using calibration curves, which show the relationship between metacognitive judgements and actual performance. They are drawn by grouping responses in particular confidence ranges and plotting assessors' mean response rate in each of these against the mean percentage correct judgement rate (Lichtenstein and Fischhoff, 1977). The resulting curve is then plotted along on a graph as shown in Figure 8. The 45 degree curve shown here represents perfect calibration, depicting individuals who have a very good appreciation of the extent of their knowledge. Judgements made with 50% confidence were correct 50% of the time, those assigned 70% confidence were correct 70% of the time and so on.

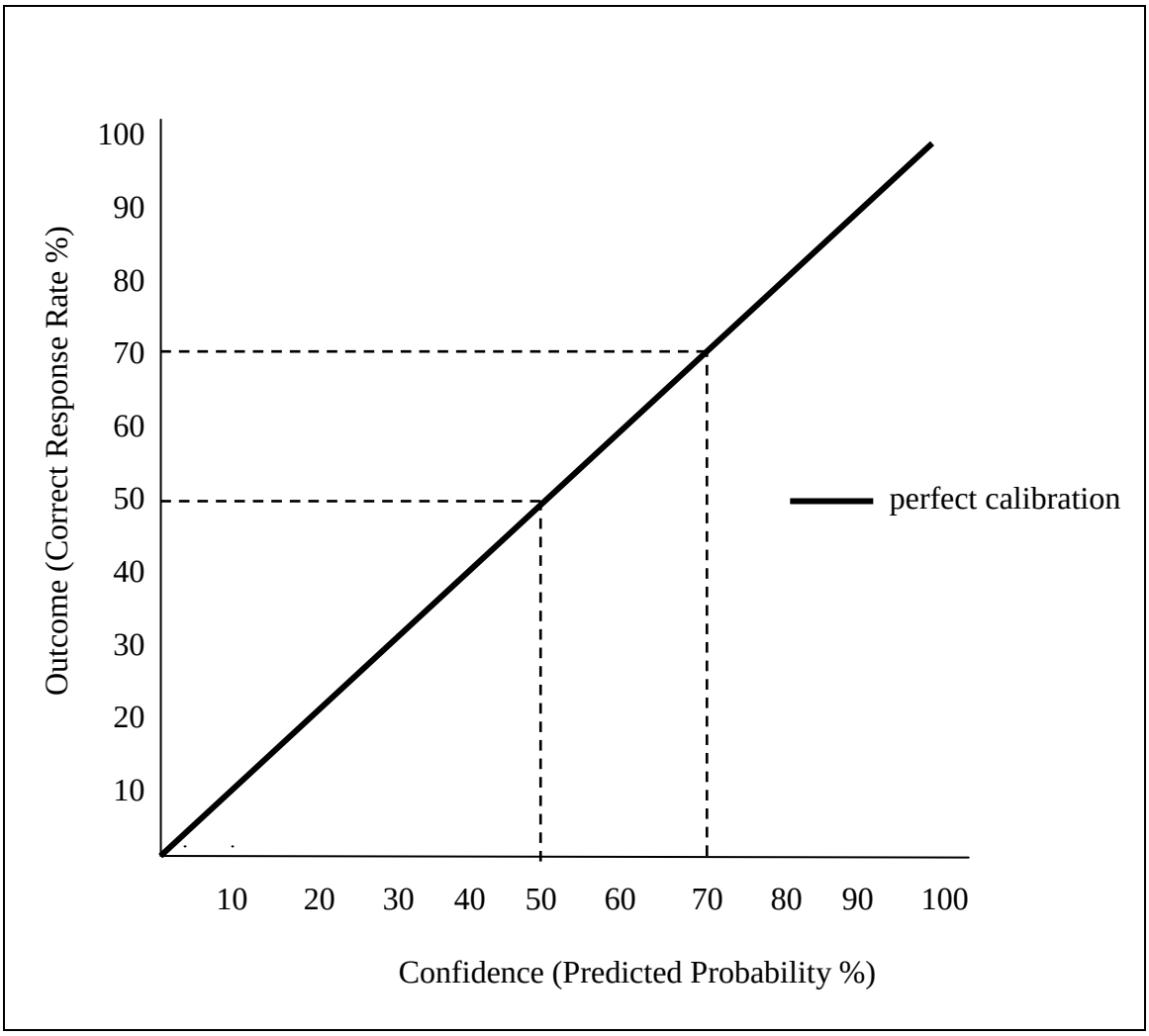


Figure 8: Calibration Curve Depicting Perfect Calibration

Thus, calibration is concerned with metacognitive monitoring in that it reflects individuals' awareness of their internal processes by indicating the accuracy of their own perception of their performance (Pieschl, 2009). Figure 9, shows how the use of calibration studies to investigate knowledge monitoring accuracy empirically, relates to the conceptual framework developed earlier.

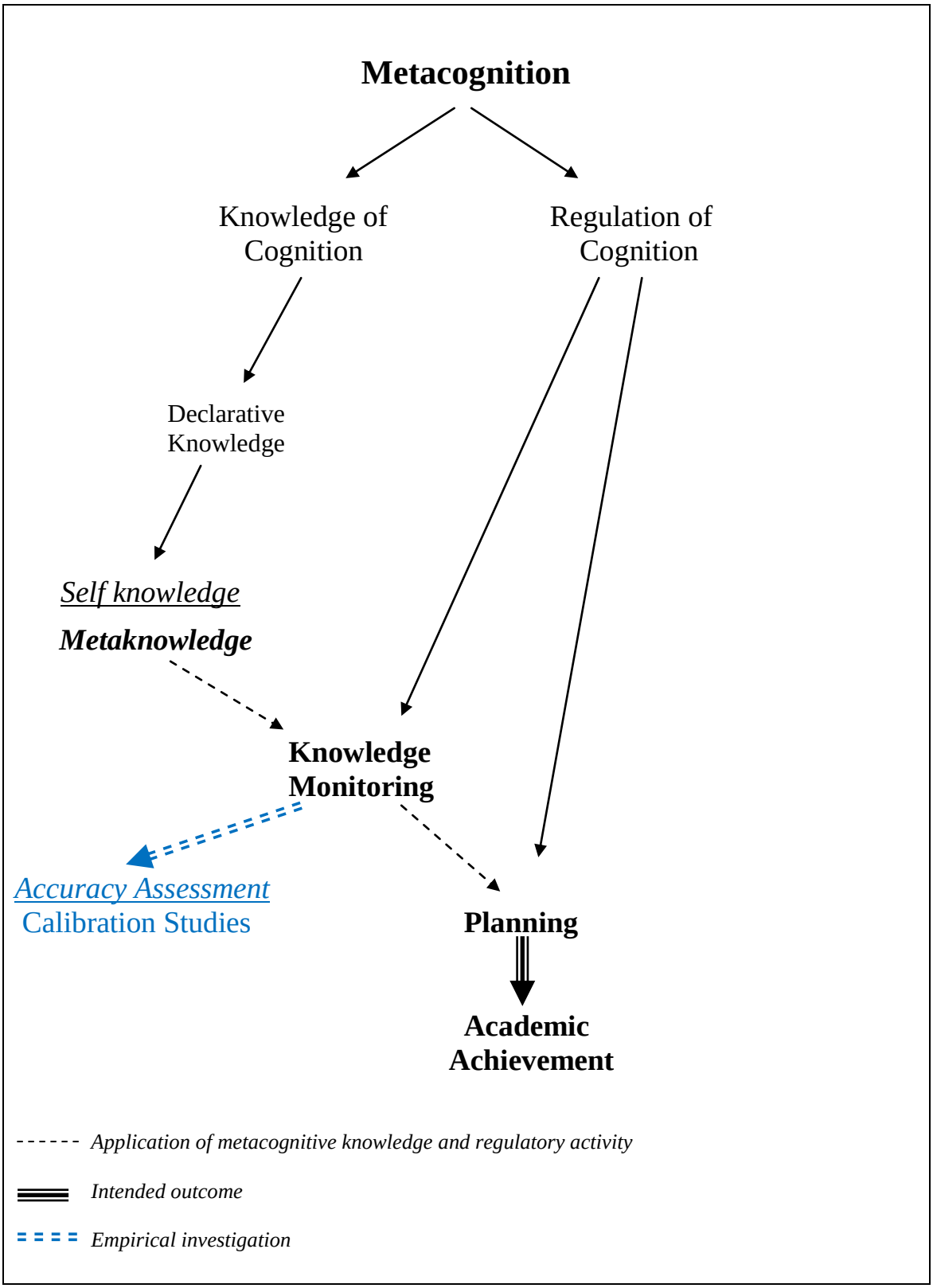


Figure 9: Calibration Studies in the Context of the Conceptual Framework

Different approaches may be adopted when assessing calibration and these can be distinguished in terms of the granularity and timing of judgements made by participants (Pieschl, 2009) as summarised in Table 1.

Table 1 - Timing and Granularity Aspects of Calibration Studies

		Timing	
		Prospective	***Retrospective***
Granularity	***Local Judgements***	Prediction of single items	Postdiction of single items
	Global Judgements	Prediction across multiple items	Postdiction across multiple items

2.7.1 Granularity of Judgements

With respect to granularity, one approach is to provide respondents with a series of questions and request a performance judgement in which they are required to provide a single (global) rating (Schraw, 2009) representing their expectation of the percentage of questions correctly answered. However, Westley (2008) argues that this method does not require participants to separately consider the extent of their knowledge of different topics and to remedy this, asked students to predict performance separately for numerous specific subject areas addressed in a test. An alternative approach, which addresses subject matter in a finer level of detail, is one in which local confidence judgements are provided in respect of each individual question. For example, as discussed above, a common approach is to provide a series of forced choice objective test questions and ask respondents to choose the option they believe to be correct and indicate their confidence level in each of those responses.

2.7.1.1 Eliciting Expressions of Confidence

Bjorkman (1994) explains that early studies of confidence judgements used verbal, rather than numerical, expressions of confidence employing categorical scales, relating to uncertainty and certainty and provoked much debate about how many categories should be used. However, things changed around 1960, when psychologists began to associate confidence with probability, which permitted the comparison of numerically expressed confidence ratings with the frequency of outcomes to investigate how realistic (Adams and Adams, 1961), or appropriate (Oskamp, 1962), confidence judgements are. As explained earlier, when using numerical expressions of confidence with multiple-choice questions, the confidence rating ranges from 100%, for answers which the respondent believes with certainty that they have answered correctly, down to a rating reflecting an answer they have only been able to guess. For example, if choosing from two alternatives, this lower level confidence rating would be 50%, reflecting the one in two probability of providing the correct answer by chance. Thus, the lower end of the confidence scale is determined by the formula 100/k, with k representing the number of alternative responses (Flannelly, 2001). These numerical confidence ratings therefore represent the observable outcomes arising from internal processing of uncertainty, in which internal cues are used to initially guide the selection of an answer and subsequently, to provide the basis for the confidence assessment (Bjorkman, 1994).

Investigating respondents' preferences between verbal and numerical expressions of confidence, Erev and Cohen (1990) found a preference for contemplating confidence in verbal terms, rather than numerically. Keren (1991) speculates that this may be due to the fact that responding quantitatively may require more effort and signify a greater commitment, as compared to a verbal response. However, using numerical expressions clarifies the interpretation of the confidence scale and improves consistency between those providing judgements, as compared with the use of verbal expressions of confidence (Adams and Adams, 1961). A study of undergraduate students by Sinkavich (1995 p.80) provides an example of the potential problems that may arise when using verbal responses. In a multiple-choice test, a five point Likert scale was used to express confidence, which ranged from +2 ('correct'), to -2 ('not correct'), with the midpoint reflecting, 'maybe it is correct; maybe it is not correct'. The first point of contention

here is the need for a response of 'not correct' in multiple choice questions. If the respondent believed that a particular option available to them *was* incorrect, then presumably they would have chosen an alternative answer. This issue also arose in an earlier study (Shaughnessy, 1979), in which a four point confidence scale was used, which included the response 'definitely incorrect'. It was reported that this was rarely used and one wonders why it was used at all, given the nature of the judgements participants were asked to make. Additionally, as Sinkavich (1995) himself acknowledged, the nature of this type of verbal scale may have resulted in inconsistent interpretation across participants and a revised version, which permitted a more stable interpretation of different confidence judgements, may have been more appropriate. Consequently, while there may be some difficulties in the use of numerical responses to assess and evaluate confidence, they may be less problematical than using verbal responses, which are subject to different, and perhaps vaguer, interpretations by respondents (Keren, 1991).

Gigerenzer (1991) argues that a problem with a numerical approach is that it requires the assignment of a confidence relating to a single event, which is then interpreted by comparing the mean of such judgements with frequencies, i.e. the percentage of questions answered correctly. He suggests that, if instead, subjects provide a global judgement of accuracy, by estimating the percentage of questions they believe they have answered correctly, poor calibration, which is suggestive of overconfidence detected in many studies using the local approach, disappears. However, Kruger and Dunning (1999) dispute this explanation and demonstrated in three different studies that overconfidence persists even where the global approach is adopted.

As illustrated earlier, using a numerical confidence scale also permits comparison of actual with expected performance at different confidence levels, which can be illustrated on calibration graphs. Where for example, a series of judgements on propositions are made with 50% confidence, an appropriate (or 'realistic') outcome would be one with a 50% correct response rate. Similarly, judgements made with 70% confidence should produce a 70% accuracy rate and 100% confidence should be accompanied by correct responses every time. Thus, the extent to which confidence can be regarded as appropriate is reflected in the degree to which it is matched by actual outcomes and

perfect calibration, representing the ideal outcome, is represented by the 45 degree line shown previously in Figure 7.

Graduation of the Confidence Scale

Fischhoff, Slovic and Lichtenstein (1977) explain that many studies using a numerical approach have employed confidence responses with 5% intervals, offering respondents options of 100%, 95%, 90% etc. to express their judgements. In other cases a broader scale has been employed, with Koku and Qureshi (2004) for example, using 20% intervals and providing respondents with only five choices, (20, 40, 60, 80 and 100%). Using fewer and pre-determined ratings may be easier to administer in that, for example, respondents could be asked to tick their chosen response to permit subsequent data analysis using optical reading equipment. However, limiting the rating scale in this manner will restrict participants' responses and Fischhoff, Slovic and Lichtenstein (1977) suggest that using a more graduated scale will allow them to better express varying degrees of uncertainty. It is also important that a consistent approach is adopted for all respondents, to avoid a confounding effect from the manner in which confidence judgements are requested. Yates et al (1989) for example, in a study comparing respondents from China, Japan and the USA, using two-choice items, asked the Japanese and USA groups to use prescribed 10% intervals, whereas the Chinese were requested to use a graduated scale in which they could indicate any confidence level between 50-100%. While they subsequently rounded Chinese judgements to the nearest 10%, this inconsistency in approach can be criticised in that it may have affected the manner in which confidence levels were provided by participants and therefore, biased the study's finding that Chinese respondents were more overconfident than those from the USA or Japan.

2.7.2 Timing of Judgements

In terms of their timing, metacognitive judgements may be elicited either pre-test or post-test (Grimes, 2002) and thus, may be either prospective or retrospective. The former requires a prediction of the outcome of a future task, whereas the latter entails postdiction, in which the self assessment of performance is provided after having

completed the task in question. Pieschl (2009) explains that studies have shown that predictions are generally less accurate than postdictions and that this has been attributed to problems related to the frame of reference of the task question. As prediction entails making judgements in advance of the task, it requires that subjects estimate the nature of the task, as well as their performance. However, with postdiction, they are already aware of the nature of the task when they make their performance judgements and as this element of uncertainty is removed, their frame of reference is clearer and judgements tend to be more accurate.

Examples of predictive and postdictive approaches on both local and global bases can be found in a study by Dunlosky and Hertzog (2000), investigating subjects' ability to recall the second word of a pair they had viewed earlier, when later prompted with the first word. Their approach is summarised in Table 2 and discussed below:

Table 2 – Examples of Granularity and Timing Differences in Metacognitive Monitoring Judgements

		Timing	
		Prospective	***Retrospective***
Granularity	***Local Judgements***	*How confident are you that in about ten minutes from now you will be able to recall the second word of the pair when prompted with the first?*	*How confident are you that the answer you just gave is correct?*
	Global Judgements	*Type any number between 0 and 100 (inclusive) that corresponds to the percentage of pairs.....that you think you will correctly recall.*	*What percentage of items did you correctly recall?*

To prompt prospective judgements, participants were initially provided with numerous word pairs (e.g. dog-spoon) and asked to provide local predictions, indicating for each pair, their confidence that they would later be able to provide the second word, having been prompted with the first. They were then each asked to make a global prediction by indicating the overall percentage of word pairs they believed they would be able to

recall. Subsequently, they were asked to make retrospective judgements by firstly providing local postdictions, indicating their confidence in the response they provided each time they were prompted with the first word of a pair. Having done so, they were each requested to make a global postdiction, by indicating the percentage of word pairs they believed they had correctly provided.

These four dimensions of calibration can be reflected in the conceptual framework developed earlier, as shown in Figure 10.

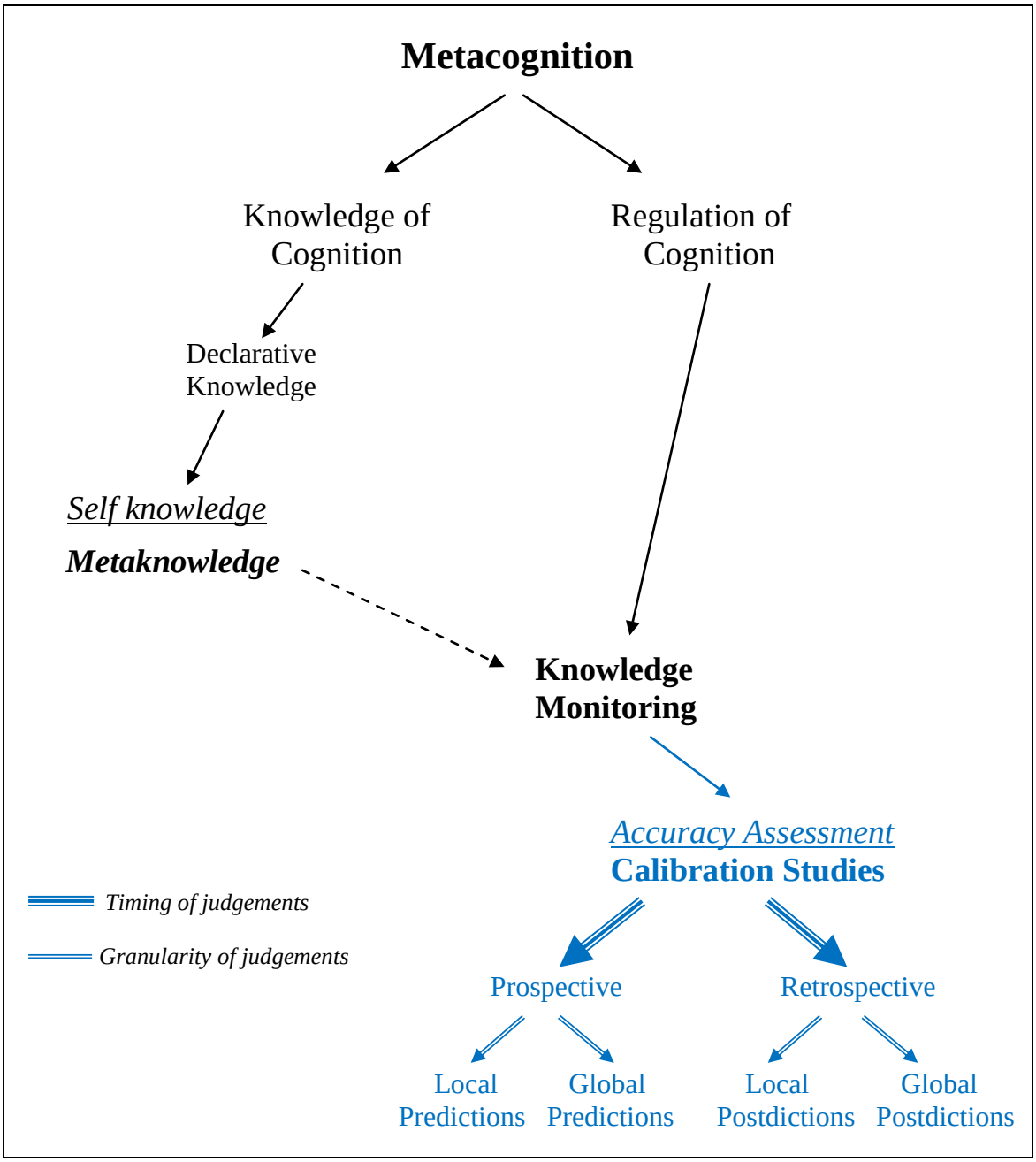


Figure 10: Dimensions of Calibration in the Context of the Conceptual Framework

A distinction can also be made between offline and online methods for investigating metacognition, where offline entails obtaining information either prior to or after the cognitive task, whereas online entails doing so while the task is being performed (Veenman, Van Hout-Wolters and Afflerbach, 2006). In Schraw's (1998) classification of components of metacognition, online assessment was referred to as monitoring while offline was classified as evaluation. In calibration studies, global judgements, whether prospective or retrospective, would lend themselves to an offline approach while for local judgements, due to their requirement to perform a number of tasks, an online approach would be more appropriate.

2.7.3 The Nature of Test Items

A fundamental issue which must be addressed in calibration studies, is the nature of the stimulus for the self-monitoring activity through which participants make meta-level judgements. As explained earlier, a common approach is to provide a series of forced choice objective questions and numerous issues must be considered when designing these.

2.7.3.1 Knowledge Domain

A commonly used method in studies of confidence is to use test items based on general knowledge. Russo and Schoemaker (1992) defend this approach by arguing that we have a responsibility to appreciate the limits of our own understanding, regardless of the subject matter. However, it has been challenged on the basis that it lacks ecological validity, which is concerned with the extent to which individuals' natural behaviour is reflected in experimental scenarios (Bem and Lord, 1979) and whether findings from social research are applicable in natural settings (Bryman, 2008). It has therefore been suggested that researchers should investigate judgements made in domains with which subjects are familiar (Dunning et al, 1990) and in natural settings (Ehrlinger et al, 2008). The ecological validity of many educationally based studies can be challenged on these grounds, in that they have been based on general knowledge or experimentally learned

knowledge, rather than richer knowledge gained over a longer period of time in a more motivational setting, such as their natural learning environment (Hacker et al, 2000). Consequently, investigation of confidence in knowledge would benefit from more studies in natural settings, based on knowledge that learners develop during the course of their studies.

2.7.3.2 Number of Distracters

A common means of studying confidence is to use two-choice questions, each comprising one correct answer and a 'distracter', in which respondents choose the answer they believe is most likely to be correct and assign a confidence level using a scale of 50-100% (Klayman et al, 1999). Thus, the two options act as a filter through which respondents communicate their responses (Shuford and Brown, 1975). Klayman et al (1999) point out that this approach has relevance to real world judgements, in that there is often a need to choose between two alternatives. However, if a respondent's knowledge state and level of uncertainty is such that they could assume more than the number of alternative responses offered, then each of these options may represent more than one knowledge state. Keren (1991) suggests that this problem can only be resolved by increasing the number of responses available. Ehrlinger et al (2008) for example, in two studies of 57 and 42 undergraduates respectively, employed a 20 item multiple choice test using five options and therefore, a confidence response scale in the range 20-100%, as did Koku and Qureshi (2004) in a study of 91 business undergraduates, using a 50 item test. However, as Keren (1991) points out, the problem with increasing the number of options is that, as respondents have different states of knowledge, and as these are unknown, it is impossible to offer an exhaustive list of options. He also argues that too many options can make it impossible for respondents to make comparisons between alternatives and assess probabilities, due to limited cognitive processing capacity. Therefore, an approach which takes these issues into account would be one in which the options were extended to some extent without over-burdening respondents with too much choice.

2.7.3.3 Number of Items

Another issue which must be addressed is the number of test items to use and previous studies have used between 20 and 300 questions (Klayman et al, 1999). Increasing the number will enhance the reliability of tests and Schraw (2009) suggests that at least six items should be used. However, a balance must be struck between this consideration and using tests which are so time consuming to complete that participants either fail to do so, or provide poorly considered responses due to boredom or a lack of concentration.

2.7.4 Operationalising Metaknowledge

When assessing metaknowledge using subjective probability, it is not possible to confirm individual judgements, except in situations where the respondent indicates 100% confidence. Hence, it is more appropriate to aggregate responses over a range of judgements and investigate the extent to which the respondent is able to assess their level of knowledge (Fischhoff and MacGregor, 1982). This requires a comparison of the aggregate probability judgements with knowledge, as demonstrated by the proportion of correct answers provided. Calibration is therefore concerned with the 'absolute fit between metacognitive judgements and performance' in a test and the two most commonly used measures are accuracy and bias, which is also known as over/underconfidence (Pieschl, 2009 p.21). Calibration accuracy is 'the mean unsigned difference between predicted and observed performances' (Pieschl, 2009 p.22) and while this definition refers to prediction, it is also suitable for the assessment of postdictive accuracy. It is a measure of the difference between confidence judgements and actual performance and while it indicates the magnitude of judgemental error, it does not provide information as to its direction, as it is unsigned. To determine this additional dimension an alternative, directional, indicator must be used, which specifies whether poorly calibrated subjects are over or under confident in their judgements. This can be determined by taking 'the difference between the mean of the probability responses and the overall proportion correct' (Lichtenstein and Fischhoff, 1977 p.161). The resulting indicator, which assesses the extent to which an individual providing

confidence judgements is over or under confident, has been referred to as the bias index and can be expressed as follows (Schraw, 2009 p.37):

$$\text{Bias index} = \frac{1}{N} \sum_{i=1}^{N} (c_i - p_i)$$

Where:

N = the total number of responses

c_i = a confidence rating

p_i = a performance score (rated as either 100% (correct) or 0% (incorrect)).

It is also known as the bias score and this term will be used in this thesis. Since it is signed, it takes into account the direction of the discrepancy between confidence judgements and actual performance (Nietfield, Cao and Osborne, 2006), with a positive score indicating overconfidence and a negative result reflecting underconfidence. Thus, it gives information in respect of both the direction and extent of the discrepancy between confidence expressed and actual performance, with its direction signifying either over or under confidence and size indicating how severe the judgemental error is (Schraw, 2009). It is the most commonly used method for operationalising overconfidence (Ehrlinger et al, 2008; Hacker et al, 2000) and has been adopted in many studies (e.g. Koku and Qureshi, 2004; Pallier, 2003; Renner and Renner, 2001; Flannelly and Flannelly, 2000; Klayman et al, 1999).

In terms of evaluating self-monitoring accuracy, it has been suggested that a bias score of less than 5% should be deemed indicative of good self-assessment (Pallier, 2003). However, while it may be applicable to overconfident individuals, using this standard more universally is problematical. It would for example, include someone with a score of say -20%, who therefore demonstrates poor self-assessment in the direction of underconfidence. Consequently, a more appropriate expression of this standard for good self-assessment would include those with a bias score of less than 5% and greater than -5%.

The calibration graphs discussed earlier can be used to supplement arithmetic indicators and have relative advantages, including providing more detail (Weingardt, Leonesio and Loftus, 1994). For example, as well as providing a visual representation of calibration, they can reflect accuracy and over/under confidence for different levels of confidence. Calibration curves are also useful when studying groups, as they provide a useful visual representation of differences between them (Pieschl, 2009). For example, in Figure 11, the fact that the calibration curve for Group One deviates from the perfect calibration line, indicates that as a group, the judges in question were poorly calibrated and its location beneath that line indicates that their poor self-assessment was in the direction of overconfidence. It also reveals that judgements made with 100% confidence produced a correct response rate of only 70% and therefore that, even when they felt they were absolutely sure about their judgements, respondents were overconfident. Group Two, since their curve falls predominantly below the prefect calibration line, also demonstrated a general tendency for overconfidence. However, since it lies above the perfect calibration line for judgements made with relatively low levels of confidence (i.e. less than 30% confidence), underconfidence is indicated for those judgements. Conversely, judgements made with confidence levels in excess of 30% resulted in overconfidence, since the curve for those falls beneath the perfect calibration line. The graph also indicates that for judgements in which they expressed 100% confidence, their accuracy was 90%, which, while also indicating overconfidence, exceeded that of Group One.

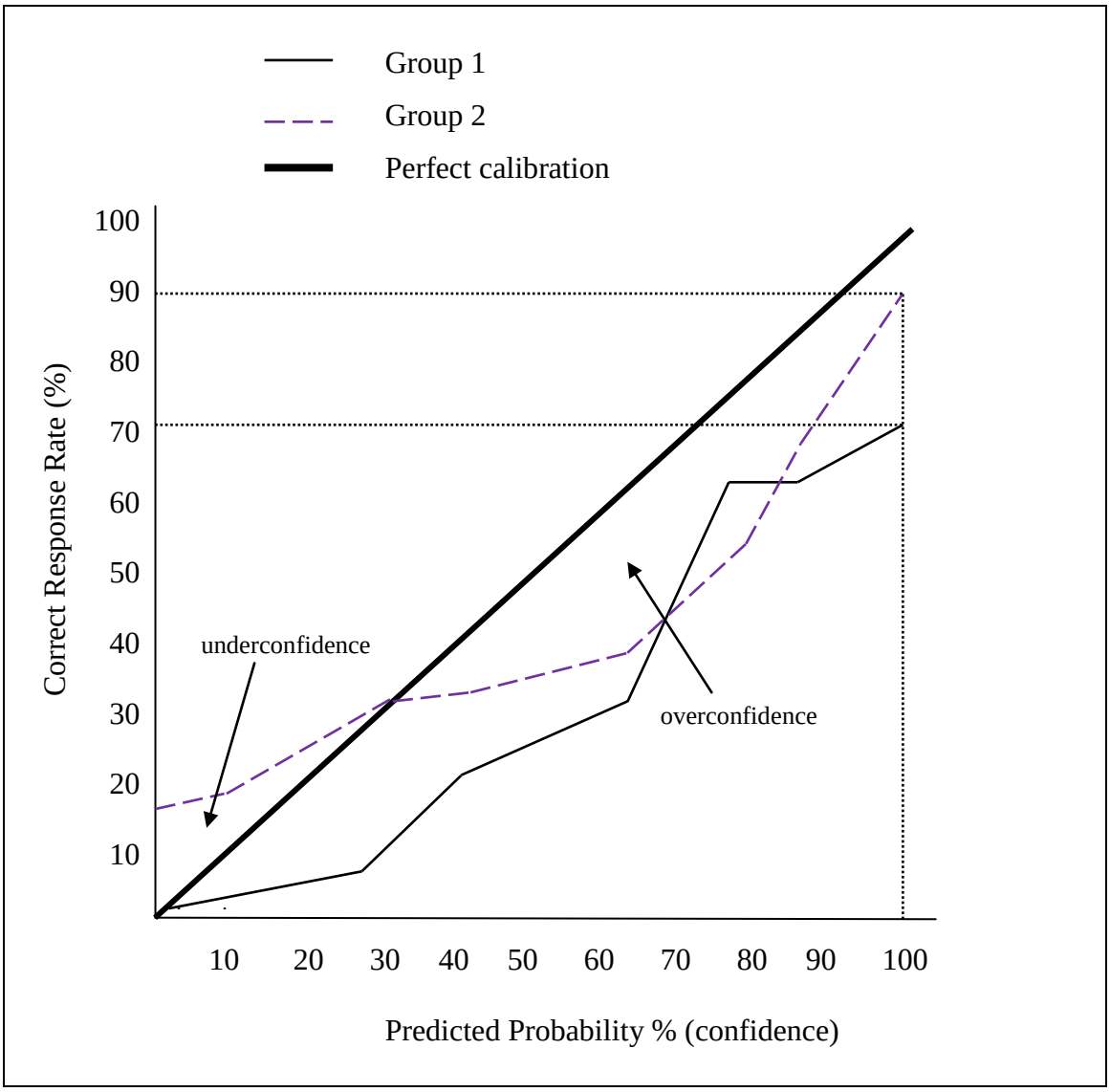


Figure 11: Calibration Curves Illustrating Group Differences

2.8 Overconfidence

Psychological research has indicated that most people are overconfident about their own ability (Camerer and Lovallo, 1999). This has been described as 'a fundamental feature of human psychology' (Bar-Tal, Sarid and Kishon-Rabin, 2001 p.77) and one of the most powerful of human tendencies, which has been linked to survival (Gilfoyle, 2000). Acker and Duck (2008) distinguish between 'referential' and 'stand-alone' overconfidence. The former relates to overconfidence about one's ability relative to

others, in which the most common view is the 'better than average' effect. This indicates that most people tend to believe they are better than average in many social, as well as intellectual, abilities (Burson, Larrick and Klayman, 2006), which as Ehrlinger et al (2008) point out, is statistically impossible. In a study of 714 engineers, for example, 42% believed that their performance was within the top 5% of the group and only one person thought s/he was below average (Zenger, 1992). Stand-alone overconfidence on the other hand, is a type of self-appraisal in which one does not relate oneself to others and miscalibration, in which judgements about knowledge are inaccurate, is an example of this.

2.8.1 Stand-Alone Overconfidence

Stand-alone overconfidence is the unjustifiable belief that one's judgements are accurate (Lichtenstein, Fischhoff and Phillips, 1982) and in studies testing self-assessment of knowledge, overconfidence is by far the most commonly reported result, with people tending to overestimate the accuracy of their knowledge (Renner and Renner, 2001). Taleb (2004) suggests that the difference between performance and self-assessment has been known since studies conducted by Meehl in 1954. However, Darwin recognised the potential problems associated with a lack of self-awareness of personal limitations as far back as the 19th century, when claiming that 'ignorance more frequently begets confidence than does knowledge' (Kennedy, Lawton and Plumlee, 2002 p.243) and thus suggesting that the less knowledge we have, the less capable we tend be in appreciating this. Bjorkman (1994 p.386) meanwhile, explains that empirical studies investigating the confidence people have in their judgements 'are as old as experimental psychology itself'. Baranski and Petrusic (1999 p.1369) trace such findings to around the turn of the 20th century. They explain that the tendency for overconfidence was highlighted over 100 years ago by Titchener, who, in a students' manual on experimental psychology published in 1905, explained that 'We feel sure, our expectation becomes conviction, long before we have the objective right to be anything more than moderately expectant'. They argue that this assertion, while prophetic, is curious, since at that time there appeared to be very little data to support it. They point out that the evidence which did exist in respect of differences between degree of certainty and accuracy suggested that it was mainly due to individual differences. Griffing (1895), for example reported that 'the

degree of confidence in the perception of intensive differences varies greatly for individuals, the proportion of wrong judgements of which observers were confident ranging from $^{1}/_{3}$ to $^{1}/_{50}$' and Fullerton and Cattell (1892) suggested that 'some observers are not confident, unless they are, in fact, right; while others are often confident when they are wrong'. Lichtenstein and Fischhoff (1977) later played an important role in reviving interest in this phenomenon, when investigating whether those who have better knowledge, also know more about how much they know (Pallier et al, 2002).

Overconfidence has been reported in many domains and while Pintrich (2002) suggests that a characteristic of experts is that they are aware of gaps in their knowledge, previous research, studying professionals in many fields, suggests that this is not necessarily so. Allwood and Granhag (1999) point out that overconfidence has been detected in studies of many professional groups, including bankers predicting stock exchange movements (Stael von Holstein, 1972), economists predicting economic downturns (Braun and Yaniv, 1992), lawyers predicting outcomes of legal cases (Loftus and Wagenaar, 1988) and FBI/CIA agents detecting lies (Ekman and O'Sullivan, 1991). Russo and Schoemaker (1992) meanwhile, investigated over 2,000 business managers, using range questions typically based on either their company, or the industry in which it operated, and reported that over 99% displayed overconfidence. For example, they found that when asked to provide 95% confidence ranges (i.e. the correct answer had a 95% chance of lying within the range provided) for questions related to their industry, computer mangers' responses fell way short of this, with only 20% of responses containing the correct answer. Similarly, advertising mangers asked to provide 50% ranges were able to provide ranges containing the correct answer for only 22% of responses. Lichtenstein, Fischhoff and Phillips (1982) examined various other similar studies, in which a 98% confidence level was used, and found that in each case, the surprise index exceeded the 2% targeted error rate (T.E.R.), with a mean for approximately 15,000 tests, of 32% (i.e. overconfidence of 30%). Plous (1993) also highlights overconfidence detected in previous research, but suggests a lower mean level of 10-20% in most studies, while Russo and Schoemaker (1992) report levels of between 30-60% in their studies of business managers. Calibration studies of surgeons have also reported overconfidence (Christensen-Szalanski and Bushyhead, 1981); while conversely, particularly good calibration has been reported for weather forecasters

(Charba and Klein, 1980). The graph in Figure 12 shows calibration curves for these studies, in which weather forecasters provided predictions of precipitation and surgeons diagnosed whether or not patients had pneumonia. The proximity of the weather forecasters' curve to the perfect calibration line is in contrast to that of the surgeons, which falls well below it, indicating unjustifiable confidence in their diagnoses.

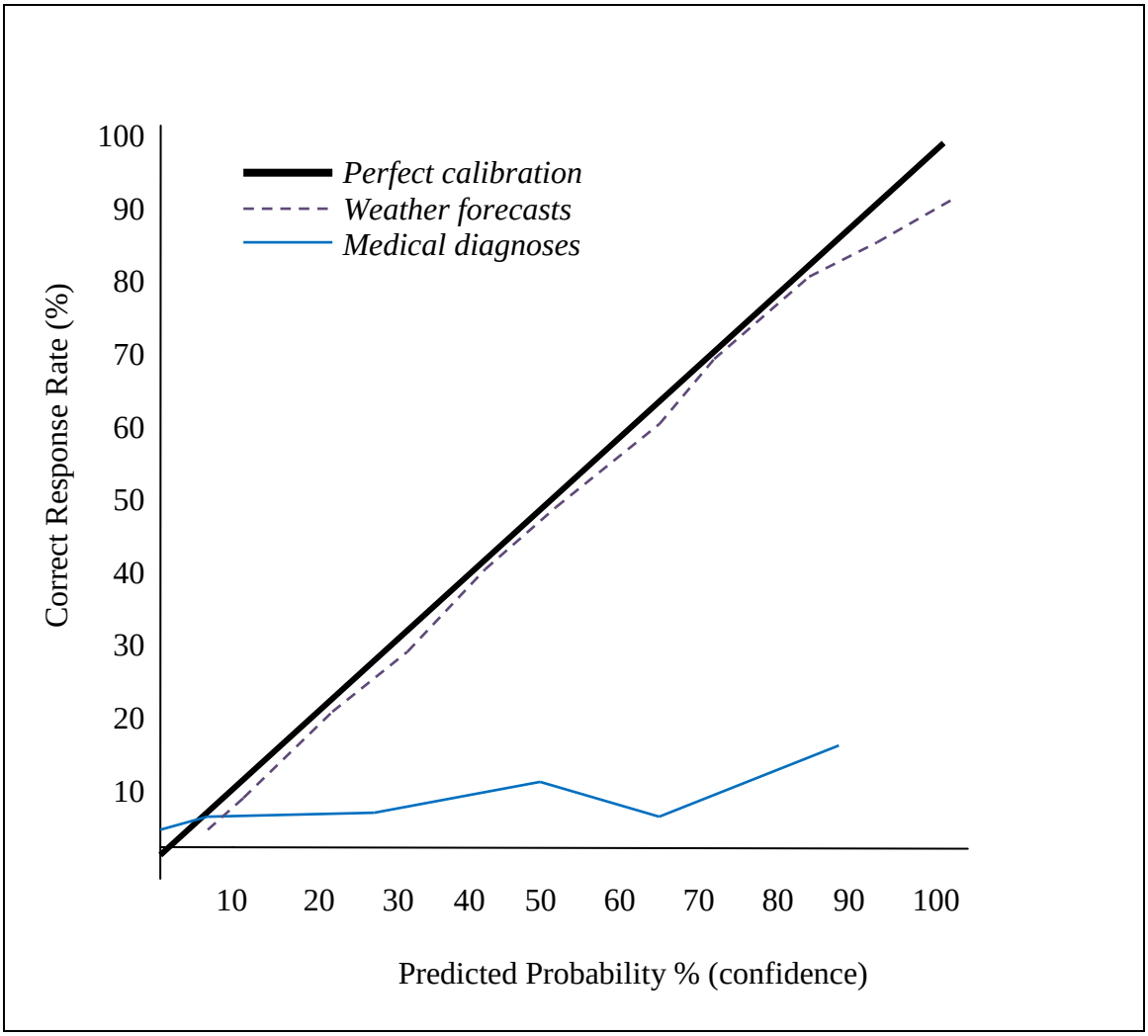


Figure 12: Calibration Curves for Weather Forecasters and Surgeons

Source: Adapted From Plous (1993)

2.8.1.1 Global vs. Local Judgements

In a calibration study using two alternative forced choice questions, which asked respondents to use both local and global judgements, Keren (1991) found that greater overconfidence resulted from the use of local judgements, evidence which seemed to support Gigerenzer's (1991) view that using global judgements reduces overconfidence. However, whereas when using local judgements subjects followed an instruction not to provide judgements below chance level, they were not given this instruction when using the global approach and more than 14% of respondents did so, by estimating a success rate of less than 50%. Eliminating these from the sample would logically increase both mean confidence, and consequently, overconfidence, for global judgements. On doing so, Keren (1991) found that overconfidence rose to a level at which it was not significantly different from that detected when using local judgements. As he points out, providing judgements below chance level suggests that such respondents either did not understand the concept, or were unable to appreciate it in the context in which they were asked to apply it and that this problem was also likely to apply to others who did not answer at below chance level. Ronis and Yates (1987) also detected respondents' use of probabilities below chance level when making local judgements, by assigning confidence levels of less than 50% for test items with two alternatives. In this case, the researchers re-interpreted such judgements by changing the choice recorded to the alternative answer and altering the confidence level to one minus that recorded by the respondent. This was based on the logic that if, for example, a respondent chose option (a) and assigned a confidence level of 30%, this suggested that their confidence in option (b) was 70% and that consequently, they actually believed that this was more likely to be the correct answer. However, as the responses provided by those participants indicate that they may not understand the probability scale, this would appear to compromise the findings from the study and Keren (1991) suggests that to address this, a screening procedure should be developed to exclude such respondents from the sample.

2.8.2 The Relationship between Confidence and Overconfidence

Klayman et al (1999) explain that the greater confidence respondents have in their judgements, the greater their overconfidence tends to be. For example, in a study of lawyers, Loftus and Wagenaar (1988) reported a positive relationship between confidence expressed and overconfidence, with highly confident predictions resulting in particularly high levels of overconfidence. However, as Dunning et al (1990) point out, if mean confidence exceeds mean accuracy and correlation between accuracy and confidence is weak, it is inevitable that highly confident judgements would also be highly overconfident. However, they argue that these results are not forced by the nature of the task or the measurement system and, where subjects had insufficient information to support high levels of confidence, they could have given lower levels.

2.8.2.1 Extreme Confidence

While it has been argued that, to some extent, poor calibration may result from subjects' poor interpretation of the confidence scale, expressions of 100% confidence present no such problem, since 100% certainty about a judgement tends to be understood by most people (Fischhoff, Slovic and Lichtenstein, 1977). Overconfidence has been found to persist in these circumstances and in a series of studies Fischhoff, Slovic and Lichtenstein (1977) detected a rate of 'false certainties' (in which 100% confidence judgements were subsequently proved to be inaccurate) of between 17% and 30% of such judgements. Dunning et al (1990) reported that this occurred for 20% of certainty responses and highlight the potential consequences of acting on judgements made with absolute certainty, which prove to be inaccurate one time in every five. They may be particularly damaging since decision makers are more likely to rely on high confidence judgements and less likely to accept disconfirmatory evidence, or take out appropriate insurance strategies in such situations. It is therefore in these circumstances, in which decisions are taken in conditions of 'inappropriate certainty', that the consequences can be most extreme (Plous, 1993 p.230). Consequently, decision makers would be well advised to think twice about things they believe they know with certainty, especially if this information is to be used to support important decisions. This view is supported by Bertrand Russell's (1951), rather blunt, observation, cited by Ehrlinger et al (2008

p.98), that ‘one of the painful things about our time is that those who feel certainty are stupid, and those with any imagination and understanding are filled with doubt and indecision’. Supportive evidence was reported in respect of a study investigating the use of absolute certainty (Fischhoff and McGregor, 1982), which found that a substantial minority of subjects tested did not give 100% confidence ratings for any of their judgements and that this group were better assessors of the extent of their knowledge.

2.8.3 The Hard-Easy Effect

Studies of bias in confidence judgements have tended to support the general tendency for gross overconfidence, with the exception of the easiest of judgements and while difficult tasks tend to result in overconfidence, easy tasks tend to produce underconfidence (Klayman et al, 1999). Thus, studies have revealed a greater tendency for overconfidence when using difficult sets of questions, where the rate of correct responses is lower, and the tendency can be reversed for easy sets of questions; a phenomenon known as ‘the hard-easy effect’ (Harvey, 1997 p.78). Positive correlation between the difficulty of judgements respondents are required to make and overconfidence has been widely reported (e.g. Koku and Qureshi, 2004; Flannelly and Flannelly, 2000; Klayman et al, 1999; Pulford and Colman, 1997; Keren, 1991; Lichtenstein, Fischhoff and Phillips, 1982; Lichtenstein and Fischhoff, 1980). It has been suggested that the hard-easy effect in fact provides supportive evidence of the difficulty people tend to have in assessing their own knowledge, since mean confidence levels are relatively insensitive to changes in the difficulty of the tasks in question (Fischhoff and MacGregor, 1982) and thus for example, when tasks becomes more difficult, confidence levels are not reduced accordingly.

However, the psychological significance of these findings has been questioned by Keren (1991), who uses the example of a two option test, in which respondents are asked to use a confidence range between 50 and 100%, to illustrate this. First of all consider at one extreme, a hypothetical situation in which questions were so difficult that respondents cannot achieve a correct response rate in excess of 50%, other than by chance. In these conditions, the mean correct response rate for a group of judges will, on chance grounds, equal 50%. Since 50% is the lowest permissible confidence response,

the mean for all such judgements cannot be lower than this figure and would therefore be equal to, or exceed, the mean correct response rate of 50%. Consequently, for extremely difficult questions, respondents could only be either overconfident or perfectly calibrated. This line of argument can also be applied, though to a lesser extent, in less extreme situations and where questions are a little less difficult, but still very challenging, where overconfidence is still more likely to occur. At the other extreme, if, hypothetically, questions were so easy that they are all correctly answered, then respondents can only be underconfident or perfectly calibrated, since their mean confidence rating across all questions cannot exceed 100%. Once again this principle can be applied to a lesser extent for judgements where accuracy is high but less than 100%. Thus, Keren (1991) argues that calibration studies incorporate a mechanism which results in underconfidence for easier test items and overconfidence for difficult judgements.

In view of this, Yates, Lee and Shinotsuka (1996) highlight the importance of using questions of equal difficulty when investigating groups using different sets of test items and explain that this is not always clearly addressed in studies. Klayman et al (1999) suggest that one method, which can be used to help prevent using questions of differing degrees of difficulty, is to focus on a very narrow knowledge domain and design questions using random sampling to determine alternative answers available to respondents. For example, a series of questions could be devised, each of which requires choosing from four different countries, the one with the highest population. In each case the available options could be determined by choosing a random sample of four from all countries in the world. However, this approach is problematical in terms of ecological validity, in that a series of judgements in such a narrow knowledge domain may not reflect the type of judgements required of individuals in more natural settings. Therefore an alternative, and more ecologically valid, approach when using different sets of questions for participant sub-groups, would be to attempt to ensure equal difficulty in their design and subsequently verify this statistically (Yates, Lee and Shinotsuka, 1996).

2.8.4 Individual Differences

Individual differences have been under-explored in respect of metacognition generally (Veenman, Van Hout-Wolters and Afflerbach, 2006) and specifically relating to self-assessment of knowledge (Ackerman, Beier and Bowen, 2002; Grimes, 2002). Previous studies which have investigated individual differences in overconfidence have focused on factors such as age, gender and nationality and Pallier et al (2002) emphasise the importance of such studies.

2.8.4.1 Age

While there is some debate as to whether metacognitive skills are transferrable to different domains, Schraw (1998) explains that many researchers believe that, while they are initially domain specific, as they develop, individuals begin to build general metacognitive skills. He suggests that this may imply that, as these skills become transferable, older and more advanced students are able to use them for a variety of tasks and thus apply them in new learning situations. Pintrich (2002) supports the view that age may be influential, arguing that some students develop metacognitive knowledge with age and experience. Veenman, Van Hout-Wolters and Afflerbach (2006) report on investigations which indicated that while metacognitive skills may generally tend to emerge at an early age and develop in future years, certain skills, such as monitoring, appear to develop later than others. Paris and Paris (2001) also explain that previous research has indicated that self-appraisal of learning improves with age and studies conducted by Fischhoff (1992) and Grimes (2002) for example, provide supportive evidence, in that each found that overconfidence tends to reduce with age.

Sternberg (1998) on the other hand, argues that while Schraw's view on the transferability of metacognitive skills may be true to some extent, it may be overoptimistic and metacognitive skills in one domain may not necessarily translate to another. He cites an example of how an individual with a good understanding of how to develop a strategy to address an issue in the physical sciences may not necessarily be able to use this competence to compose a literary piece of work. Pallier (2003) found evidence which contradicts the view that metacognitive skill is better for older people, when reporting a reduction in monitoring accuracy with age. Fitzgerald, White and

Gruppen (2003) meanwhile, suggest that while age may be influential to some extent, development of self-assessment ability may not be progressive throughout life span, explaining that one school of thought suggests that it is largely learned in childhood, and by adulthood is fixed. They suggest that the lack of evidence of development of self-assessment skills in a study of medical students may either support this view or alternatively, may result from the lack of emphasis on the development of self-assessment skills in education.

2.8.4.2 Gender

Although the tendency has been demonstrated in both males and females, it has been reported that overconfidence tends to be more evident in males (Acker and Duck, 2008; Pallier, 2003; Barber and Odean, 2001; Lundeberg, Fox, and Punccohar, 1994). Beyer (1990 p.960), who suggests that research on the accuracy of self-monitoring should devote more attention to gender differences, explains that females have been associated with lower expectations of success in many domains. She argues that this may be attributable to a tendency for 'self-derogatory' bias, through which ability is underestimated. This is in contrast to males, for whom the tendency for 'self-enhancing' bias is more prevalent. This tendency for a lack of confidence in females has been suggested as a factor which may reduce their persistence in higher education and even when outperforming males, they have a tendency to underestimate their capability (Lundeberg, Fox, and Punccohar, 1994). However, it has been suggested that gender differences in some cases arise not because females lack confidence, but rather that males display too much, a tendency which has been referred to as 'male answer syndrome' (Lundeberg, Fox, and Punccohar 1994, p.120). Pallier (2003) for example, in a study of 303 subjects drawn from the general population in Sydney Australia, found that males displayed significantly higher levels of confidence, which were not accompanied by better test performance, and consequently, resulted in significantly more overconfidence than for females. Dunning et al (2003) similarly reported on a study using a series of scientific reasoning questions in respect of which, despite there being no significant difference in actual results, males provided higher estimations of their performance. It has been argued that this may result from the greater influence on males of self serving attribution bias, in which too much personal credit is assumed for

one's successes (Beyer, 1990; Meehan and Overton, 1986; Deaux and Farris, 1977). Overconfidence has also been associated with risky behaviour and risk taking generally tends to be more common in males than females (Nicholson and Willman, 2001). In terms of how gender and age may interact, Pallier (2003) found that males were more overconfident regardless of their age. Pressley et al (1987) meanwhile also found that at a young age, girls are better able to monitor their knowledge than boys. They also reported that older girls outperformed boys of the same age to an even greater extent, as their monitoring ability increased with age, whereas the boys' did not.

However, greater overconfidence in males is not a universal conclusion and Bromily and Curley (1992) suggest that in certain contexts, the reverse may true. Pallier (2003) for example, highlights the potential importance of the knowledge domain, explaining that gender stereotypes suggest that females tend to have more confidence in humanities subject areas and males in science based subjects. Addressing ecological validity in the context of gender studies of confidence ratings, Lundeberg, Fox, and Punccohar (1994) argue that, while there have been consistent findings indicating the tendency for females to display lower levels of confidence than males in studies of general knowledge, there is less known about such differences in the context of more natural settings, such as answering tests or examination questions, as few researchers have addressed this. More recently, Pallier (2003) also argues that gender differences in findings related to overconfidence appear to be under-investigated and this is therefore an issue which merits further study.

2.8.4.3 Nationality

Previous studies have detected particularly high levels of overconfidence for Asians, with Yates, Lee and Bush (1997), for example, finding that they were more overconfident than Westerners. In another study, based on undergraduate students' confidence in self-estimates of their end of year examination marks, Acker and Duck (2008 p.1817) investigated the difference between British and East Asian students (classified as originating from, China, Hong Kong, Malaysia or Taiwan) and found that Asians displayed greater overconfidence than their British counterparts. They suggest that this may be due to over-optimism, or the fact that according to the Risk Avoidance

Index, which measures 'the level of tolerance for uncertainty and ambiguity within a society', China and Hong Kong are among the less risk averse cultures, which may indicate a greater tendency to use high levels of confidence in judgements, rather than expressing greater uncertainty. Extreme overconfidence has been reported more for subjects from a Chinese cultural background than any other (Yates, Lee and Shinotsuka, 1996). Culpepper, Zhao and Lowery (2002) explain that, while there is a general tendency for overconfidence regardless of nationality, various studies have indicated that for Chinese subjects this is relatively extreme. They suggest that extreme overconfidence for Chinese subjects may be attributable to a tendency for less debate between opposing views in Chinese culture and that this may account for a tendency to use extreme responses. However, they also highlight an opposing view, suggesting that Chinese respondents may be more likely to display modesty than overconfidence, since this, rather than assertiveness and strong opinions, is more valued in China, because Asian cultures tend to be more collectivist than individualist and this self-effacement serves to promote harmony between individuals (Yates, Lee and Shinotsuka, 1996). These arguments suggest that there may be two opposing influences on Chinese subjects when asked to express confidence in judgements and Culpepper, Zhao and Lowery (2002) explain that there is empirical evidence supporting each of these. They help to reconcile these findings by explaining that they may differ according to the nature of the task under investigation. Idiographic tasks, which emphasise the individual, may be more likely to produce a modest response around the mid-range. Chen, Lee and Stevenson (1995) found evidence of this in a study investigating cross cultural differences in the use of response ranges in surveys, by comparing Asian and North American students. Their questions included many of an idiographic nature and they reported that responses from Chinese participants were more biased towards the mid-point than North Americans. However, for more nomothetic enquiries and tests based on general knowledge or knowledge of specific facts, the tendency for Chinese to consider less contrary evidence may be more likely to result in extreme responses and increase the potential for overconfidence.

In terms of the influence of other individual characteristics, Acker and Duck (2008) also investigated interaction between nationality and gender and found no significant difference between Asian males and females in terms overconfidence. However, as their

sample included only 18 Asians and 68 British students, this limits the generalisability of their findings and this issue would benefit from further investigation with larger samples.

2.8.5 Overconfidence in Educational Environments

Lundeberg, Fox and Punccohar (1994), explain that appreciating what is known and what is not, has important implications for study behaviour. When assigned a learning task, students must accurately assess the difficulty of the material they must learn, their own understanding and the task requirement (Isaacson and Fujita, 2006). The second of these requires that they are able to assess their own knowledge accurately as this will guide them in devoting more time to areas in which knowledge is judged to be poor, or curtailing study when it is perceived to be satisfactory (Smith, Shields and Washburn, 2003). However, as Kennedy, Lawton and Plumlee (2002) point out, the tendency for students to display overconfidence in their knowledge has been reported in numerous studies (e.gs. Koku and Qureshi, 2004; Dunning et al, 2003; Pallier, 2003; Grimes, 2002; Pallier et al, 2002; Renner and Renner, 2001; Ramnarayan, Strohschneider and Schaub, 1997; Yates, Lee and Bush, 1997; Alicke et al, 1995; Brown and Gallagher, 1992; Oskamp, 1962).

Hacker, Bol and Bahbahani (2008) refer to findings suggesting that calibration is relatively stable over time and across tasks, suggesting that this may imply that it is not controllable but a permanent personal trait. However, in a study of business and sociology students at two universities in the USA, Kennedy, Lawton and Plumlee (2002) reported that overconfidence declined with the length of time students had spent on their courses. They suggest that this indicates that the phenomenon was not due to personality traits alone and that the educational experience had indicated the limitations in their own understanding and highlighted how much more they had to learn. This explanation contradicts Russo and Schoemaker's (1992) suggestion that metaknowledge tends not to be developed during formal education.

It has been suggested that epistemological beliefs are an important influence on cognitive monitoring (Tynjala, Helle and Murtonen, 2001) and in this respect Jehng, Johnson and Anderson (1993) found that those studying at postgraduate level are more

likely than undergraduates to recognise that knowledge is uncertain. Differences between disciplines in education and the tendency for them to attract different personality types may also mean students in one discipline having different beliefs, or behaving differently, from those in another (Koku and Qureshi, 2004). Consequently, Jehng, Johnson and Anderson (1993) also investigated differences between students from four academic fields of study: engineering, arts and humanities, social science and business, in terms of their recognition of the uncertainty of knowledge. For comparative purposes they classified engineering and business as 'hard' fields and arts and humanities and social sciences as 'soft' and found that those in 'hard' fields had less tendency to believe that knowledge is uncertain. However, their classification of business alongside engineering as a 'hard' field can be challenged in that it was based on an assumption that, 'typical problems presented in engineering, natural science, and business courses contain enough information to reach a particular solution' (Jehng, Johnson and Anderson, 1993 p.33). In the case of business, this is often not the case, since information supporting many business decisions is insufficient to reach clear and definitive conclusions.

When investigating subject specialism differences in a meta-analysis of previous self-evaluation studies in higher education, Falchikov and Boud (1989) found that students from the natural sciences provided more accurate assessments of test performance than those studying in the social sciences. Ackerman, Beier and Bowen (2002) meanwhile found that those who had specialised in business at university displayed lower self-assessment accuracy than those studying the physical sciences, social sciences and humanities. Koku and Qureshi (2004) however, highlighted a lack of studies exploring the relationship between overconfidence and examination performance for business students and investigated 91 undergraduates specialising in business. Using a research instrument based on a multiple choice examination, they found that the group as a whole displayed overconfidence. They also reported evidence of the hard-easy effect in that difficult questions produced extreme overconfidence, with a bias score of +36% (i.e. a mean confidence score of 73% vs. an accuracy rate of 37%), intermediate questions lower overconfidence of +7% (77% vs. 70%) and easy questions slight underconfidence, with a bias score of -2% (84% vs. 86%). Interestingly, they express

surprise at the underconfidence detected for easy questions, despite evidence of the hard-easy effect reported in previous studies.

2.8.6 Overconfidence and Competence

Dunning et al (2003) explain that there is much evidence that less competent performers tend to have greater difficulty in making metacognitive judgements than the more competent. The inability of the incompetent to appreciate their shortcomings has been reported in numerous domains, such as social competence, physical reaction times and comprehension of written text (Kruger and Dunning, 1999). In respect of metaknowledge, Pintrich (2002) argues that experts tend to have a good idea of what they do not know. However, contradictory evidence is provided by the numerous studies that have detected overconfidence in professionals in many domains and which therefore suggest that either most of those tested lack expertise, or that well developed metaknowledge is not necessarily associated with competence.

In terms of its association with academic achievement, while most empirical evidence shows low to medium levels of calibration accuracy, it indicates that this is particularly so for poor learners (Pieschl, 2009) and previous studies have shown a positive relationship between accurate self-appraisal and academic achievement (Paris and Paris, 2001). Tobias and Everson (2002 p.21), meanwhile, when discussing a programme of research investigating knowledge monitoring ability conducted over more than 10 years, conclude that their findings suggest that the ability to distinguish between what is known and what is not known ‘is an important ingredient for success in all academic settings’.

2.8.7 The Association between Knowledge and Overconfidence

Since multiple choice tests used in calibration studies can be used to derive indicators of both knowledge (i.e. the proportion of correct answers) and metaknowledge (i.e. the bias score), they provide researchers with the means of investigating whether those who know more, also know more about how much they know. For example, in an important early study addressing this issue Lichtenstein and Fischhoff (1977) investigated 120 university students, who responded to a call for paid volunteers, using a test based on

general knowledge. They split these into three sub-groups according to their knowledge scores on the test and results indicated that the higher performers had a mean bias score of 5% (76% mean confidence vs. 71% accuracy), the middle group 7% (71% vs. 64%) and the worst performers 15% (71% vs. 56%). Thus while each group tended to display overconfidence, the more they knew the better calibrated they were. Each respondent answered 75 test items drawn from a pool of 150 general knowledge questions and while this meant that they completed different sets of questions, task difficulty did not represent a potential confounding effect, since individual differences were not investigated. However, since the test was conducted in a two hour session, in which participants also took part in a number of other judgemental activities, it may have been compromised by them providing ill-considered responses due to poor concentration or boredom. This may have been compounded by a lack of motivation when completing the tests due to their participation being prompted by the inducement of a cash payment. Additionally, since general knowledge questions were used, the ecological validity of the study can be challenged on the grounds that the results may not have been replicated in a more natural setting.

However, similar findings have been detected by others in more natural settings. Sinkavich (1995) for example, found that those performing poorly in university educational psychology examinations were less able to appreciate which question they answered correctly and which incorrectly, when compared with those who performed well. However, in this case, verbal, rather than numerical, expressions of confidence were used and these may be more likely to be inconsistently interpreted by participants. Flannelly and Flannelly (2000) on the other hand, used numerical expressions of confidence in a study of undergraduate nursing students' performance in tests based on their study programme. They also reported a negative relationship between test performance and overconfidence bias and suggested that overconfidence poses a problem for learning. Renner and Renner (2001) also found a negative association between knowledge and overconfidence in an experiment involving psychology students, in which participants completed a series of weekly subject-related tests over a ten week period. These were four-option multiple choice tests in which respondents provided confidence estimates between 25% and 100% for each of 10 questions and were given feedback on their performance. Changes over time were assessed by

comparing the means for the first two tests with those for the last two. This indicated that at the start of the experiment, participants were generally overconfident in their knowledge, with a mean bias score of +6%, but by the end, were underconfident, with a bias score of -14%. The study also revealed that this group's performance on the tests was significantly better at the end of the experiment than at the start, while for a control group, which had not provided confidence estimates, performance had declined. Thus, the researchers argued that the reduction in the bias score was accompanied by better performance in the tests. In this case while overconfidence was reduced, participants were not better calibrated, since at the end of the experiment they were underconfident to a greater extent than their initial overconfidence (-14% vs. 6%). However, Renner and Renner (2001 p.31) argue that underconfidence may be the preferable error, since it is less likely than overconfidence to result in inadequate learning in preparation for tests and support this view with a quote from one participant who remarked, 'I like being underconfident, it makes me study more'.

Everson and Tobias (1998) also reported that less competent participants were less able to appreciate their own level of knowledge in a study of university students at the start of their programme, in which vocabulary knowledge was assessed. However, the study did not address subject knowledge specific to their programme of study, nor did it consider a wide range of competence, since participants were from a group identified as at risk of performing poorly in their studies. In another study, based on an examination taken by students with a wider ability range, Ehrlinger et al (2008) grouped their 124 participants by competence displayed in the test and found that the bottom quartile and middle 50% displayed significant overconfidence, while the upper quartile estimated their performance much more accurately and displayed underconfidence to a small extent. Meanwhile in the study of business students discussed earlier, Koku and Qureshi (2004) also reported a significant negative relationship between overconfidence and knowledge for business undergraduates. They classified participants as high, average or low performers based on their test score in the multiple choice examination on which the research instrument was based, in the same manner as Lichtenstein and Fischhoff (1977), with the top third classed as high performing the middle third as average and the remainder as low performing. They found greater differences between these groups than Lichtenstein and Fischhoff (1977), with high performers only slightly overconfident

with a bias score of 3% (83% mean confidence vs. 80% accuracy), average performers moderately overconfident with a bias score of 10% (78% vs. 68%) and low performers extremely overconfident, with a bias score of 22% (73% vs. 51%). While these studies detected relatively low levels of overconfidence for higher performers, studies of university students' global postdictions of multiple-choice examination performance on introductory educational psychology courses (Hacker et al, 2000; Hacker, Bol and Bahbahani, 2008), reported underconfidence for the more knowledgeable. They also indicated that those who performed best were better calibrated and it was argued that these findings suggest that a high degree of accuracy in calibration may only occur where an individual has very good knowledge in the domain being tested. However, while Hacker, Bol and Bahbahani (2008) explained that their results should be viewed cautiously due to the relatively small number of participants, they did not discuss the implications of the gender split in the sample. Participants were predominantly female (80%) and consequently, it may be that their findings were gender biased. Kruger and Dunning (1999 p.1121), who also detected that those who displayed least competence were most likely to overestimate it, argue that the skills necessary for competence are often the same as those required to evaluate it. They conclude that this lack of metacognitive skills for accurate self-assessment represents a 'dual burden' for some in that they not only make incorrect choices but 'their incompetence robs them of the ability to realize it'.

However, while many studies have indicated a negative association between overconfidence and knowledge, it has been argued that the method used in such studies has biased findings. This argument suggests that those with differing levels of competence may not differ in their self-assessment ability, but that people of different skill levels find it equally difficult and this, combined with problems in methods adopted in research studies, produces findings suggesting the less competent are poorer at self-assessment (Ehrlinger et al, 2008). As Klayman et al (1999) point out, the bias score is determined by comparing confidence and accuracy and therefore overconfidence and accuracy will be negatively correlated when using the same set of questions to measure each. Therefore, those displaying poor knowledge in the type of test commonly used in calibration studies are more likely to show overconfidence than those showing good knowledge, because it is more difficult for them to underestimate

their performance, since it is so poor (Kruger and Dunning, 1999). Similarly those displaying high knowledge levels are more likely to display underconfidence, since is more difficult to overestimate such high performance. To take an extreme example, it would be impossible for someone answering 100% of test items correctly to display overconfidence, since they would be unable to express a mean confidence level in excess of 100% for those judgements. However, when investigating this issue, Kruger and Dunning (1999) found that with training, the self-assessment ability of less knowledgeable participants was enhanced and they were able to provide better self-assessment, even where test scores were low. They argue that this suggests that, while it may play some part, overconfidence in the less knowledgeable does not arise solely due to the lower likelihood of underestimating low scores. However, since the problem identified occurs through the use of the same instrument to operationalise both competence and calibration, it could be addressed in an educational context by using a separate instrument to independently assess academic competence. For example, assessment marks during participants' study programme could be used for this purpose. Since performance indicators used when investigating the association between calibration and performance have tended to be restricted to objective, multiple choice tests, little is known about how it relates to broader indicators of academic performance (Nietfield, Cao and Osborne, 2006). The approach suggested here would address this by permitting the investigation of the association between overconfidence and indicators of student performance which impact on the overall grade awarded in their study programme.

2.9 Reasons for Overconfidence

The potential adverse consequences of overconfidence raise questions about its possible causes and whether interventions can be initiated to assist in moderating it. Reasons for overconfidence detected in research studies can be categorised as cognitive explanations, motivational explanations or flaws in testing.

2.9.1 Cognitive Reasons

Overconfidence may arise through the use of heuristics when making judgements. These are mental shortcuts which individuals take when making judgements, in order to

simplify the task. However, while they may be advantageous in reducing the cognitive effort required when making judgements, they can also cause problems. The following heuristics have been identified as potential causes of overconfidence (Russo and Schoemaker, 1992).

2.9.1.1 Availability Bias

Availability bias arises due to the difficulty people have in picturing all the ways in which events could possibly unfold (Tversky and Kahneman, 1982). This results in a tendency to over-rely on evidence which is most available or more easily accessed, perhaps based on recent experiences for example, and to give insufficient consideration to other sources. Thus, unwarranted confidence in judgements may arise through consideration of over-restricted sources of relevant information. For example, when mentally searching for sources of information to inform a response in an educational test, students may over-rely on recently learned material, or issues which have received more prominence in previous teaching sessions.

2.9.1.2 Confirmation Bias

Confirmation bias arises due to the tendency to seek support for expectations rather than looking for disconfirmatory evidence (Russo and Schoemaker, 1992). Selectively adopting or ignoring information, depending on whether it fits expectations, can result in failure to consider alternative possibilities and thus place too much confidence in the initial perception. For example, a student may feel that s/he knows an answer and fail to subject this belief to sufficient critical examination by considering alternatives (Koku and Qureshi, 2004). This bias can work in conjunction with availability bias in that the availability of information may influence initial expectations, which the student subsequently fails to assess critically enough when evaluating evidence to inform his/her judgement.

2.9.1.3 Anchoring Bias

Anchoring bias results from a reluctance to shift sufficiently our belief from one particular notion on which we have become ‘anchored’. In a seminal paper on judgement under uncertainty and the use of heuristics, Tversky and Kahneman (1974)

explain that when making judgements, individuals often use a start point (anchor) and adjust from that in making their judgement. Problems tend to occur through either choosing an inappropriate anchor or failing to adjust sufficiently from that starting point to take other relevant information into account and Tversky and Kahneman (1974) suggest that this can account for overconfidence.

Keren (1991) argues that anchoring may help to explain the hard-easy effect in calibration studies. He suggests that when uncertain about the answer to a test item, rather than indicating a confidence judgement near chance level, respondents may use an anchor of probability based on moderate difficulty in the middle of the confidence rating range available to them. For example, as illustrated in Figure 13, for two-choice questions this would be around 75% % [50 + (100-50)/2] and for four-option questions, somewhere around 60-65% [25 + (100-25)/2].

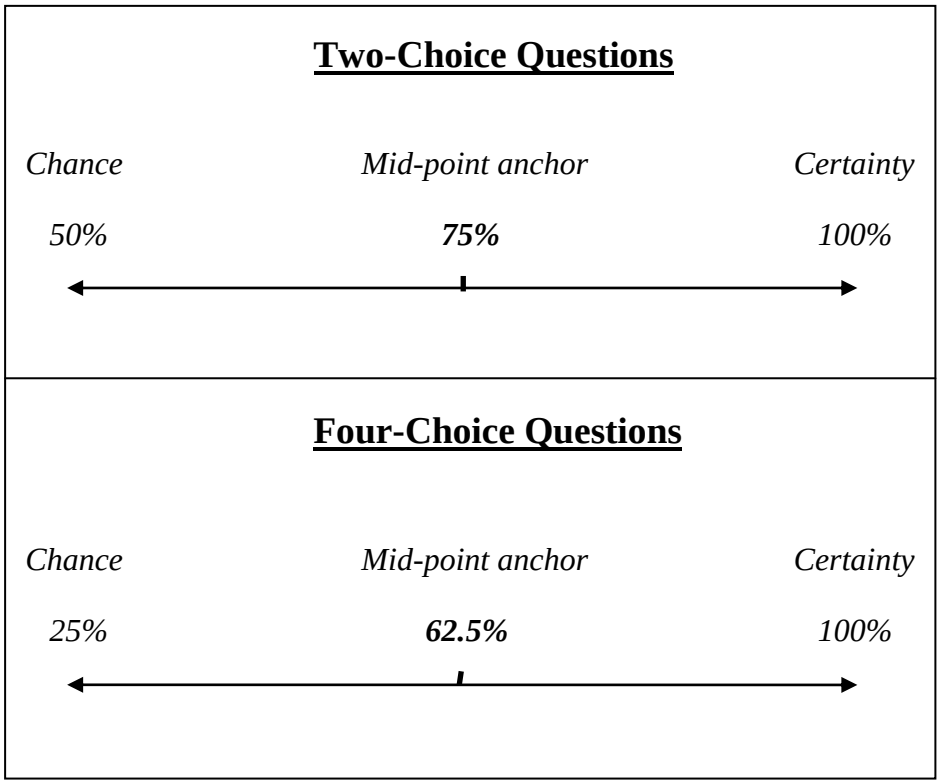


Figure13: Location of the Mid-Point Anchor in Multiple-Choice Questions

While respondents make adjustments from this initial anchor, to account for questions which are perceived as being either particularly easy or difficult, these are often insufficient. This can result in confidence estimates from those who are uncertain in their judgements being biased towards this central reference point, causing 'response contraction bias' (Poulton, 1994 p.13). Figure 14 illustrates this effect using two-choice questions for illustrative purposes. It shows how it would tend to result in a stimulus which is higher than the central reference point (i.e. from easy questions) being underestimated, resulting in underconfidence, while stimuli below the reference point (i.e. difficult questions) would be overestimated, and produce overconfidence.

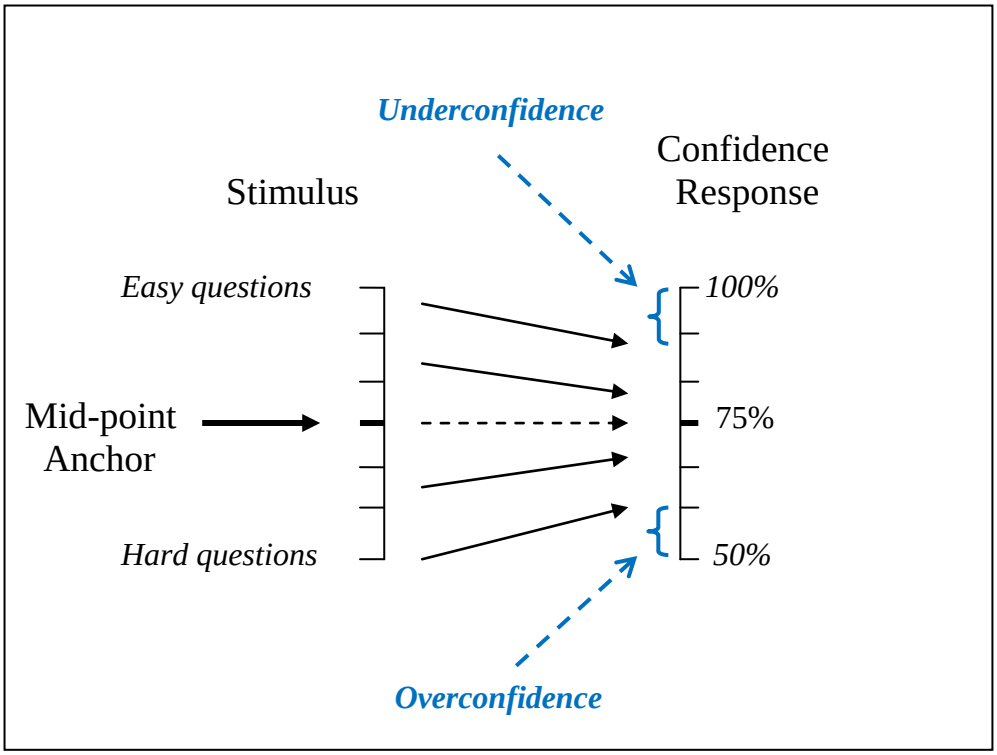


Figure 14: Response Contraction Bias

Source: Adapted from Poulton (1994)

Respondents behaving in this manner thus tend to play safe, by contracting their confidence judgements to the mid-point (Poulton, 1994). Consequently, to assist in preventing this tendency, they should be encouraged to use the full confidence range to reflect different states of uncertainty more appropriately.

2.9.2 Motivational Explanations

2.9.2.1 Social Utility

It has been suggested that motivational factors may be a contributory factor to the overconfidence detected in research studies, in that respondents may feel they should appear knowledgeable and confident, since in many situations competence can be difficult to assess and confidence may therefore be rewarded instead (Klayman et al, 1999 p.243). This suggests that respondents may feel motivated to express greater confidence than is warranted if they feel it is to their advantage to do so (Wright and Wishuda, 1982). Fischhoff and MacGregor (1982) argue that this implies that findings from studies investigating confidence of those who are consulted, and paid, on the basis the confidence they project (e.g. stock market advisors and management consultants) may be ambiguous, since they may feel motivated to express confidence in their judgements, whether warranted or not. People like to see themselves as knowledgeable (Klayman et al, 1999) and the less competent, in particular, may display overconfidence, not because they are unable to accurately self-assess, but rather they are unwilling to do so and admit to their incompetence, whereas those performing well have less incentive to do so and indeed may even display false modesty (Ehrlinger et al, 2008).

Camerer and Hogarth (1999) analysed 74 previous studies and concluded that monetary rewards for good calibration can act as an effective means of addressing this issue, as respondents offered this competing motivation tend to shift responses, from favourable self-presentation to a more reasonable response. However, Ehrlinger et al (2008) found evidence to the contrary in a study in which they investigated the impact of motivation on the less competent, by offering monetary rewards to one group for accurate self-assessment and comparing their responses with a control group, for whom there was no such incentive. They found that with a monetary incentive, less competent respondents actually displayed more overconfidence, rather than less as would have been predicted by a motivational explanation. They also investigated the impact of social incentives, by requiring respondents to explain their responses to the researcher, on the grounds that people tend to make more considered judgements when they are accountable for them.

Here they found similar results to those obtained with financial incentives and therefore concluded that for the group in question, overconfidence in the less competent was not attributable to motivational explanations.

2.9.2.2 Wishful Thinking

Another influence on overconfidence may be the extent to which a respondent is involved in the outcomes of the judgement and the wishful thinking that may consequently result (Keren, 1991). Babad (1987) detected this in a study of predictions of football results, where respondents were more likely to predict a win for a team to which they had some affiliation. In a study of psychology students, Hacker, Bol and Bahbahani (2008) also found some evidence of this, albeit from a very small percentage of their sample, with participants assigning confidence levels based on their desire for high scores.

2.9.3 Flaws in Testing

2.9.3.1 Question Difficulty

While there was general agreement from the 1970s until the early 1990s that individuals were overconfident about the state of their own knowledge, a view emerged in the 1990s which suggested that this finding may have been influenced by the nature of the questions used in testing (Klayman et al, 1999). It has been claimed for example, that overconfidence detected in some studies may not be generalisable, since it may have arisen due to flaws in the way in which test questions were selected, with researchers using particularly difficult questions (Juslin, 1994). Contrary questions are those for which the correct answer contradicts the information subjects tend to use to guide them in making a judgement (Klayman et al, 1999) and it has been argued that the use of such misleading test items introduces an experimental bias, which tends to produce overconfidence (Klayman et al, 1999; Keren, 1991). However, it can be difficult to take this into account when designing tests since question difficulty depends on the respondents and may differ for different sub-groups (Keren 1997) and for certain individuals. Klayman et al (1999) also argue that in natural settings, people are required

to make some judgements which may be more difficult than others and an understanding of how they respond to such difficult questions is therefore important.

2.9.3.2 Difficulty in Expressing Confidence Levels

Koku and Qureshi (2004) suggest that respondents' difficulty in translating subjective judgements into quantitative measures may have contributed to overconfidence detected in research studies. They argue however, that it cannot be solely due to this on the grounds that one would expect to find underconfident as well as overconfident individuals and previous studies have generally reported overconfidence. However, this argument can be challenged, in that it does not take into account the fact that some studies have found evidence of underconfidence. For example, it has been detected for easy questions (i.e. the 'hard-easy' effect) and, as discussed above, response contraction bias, in which respondents misuse the confidence scale, has been suggested as a possible cause. Consequently, while numerical expressions of confidence may be less subjective than verbal responses, it is important that studies adopting this approach take steps to clarify for respondents the nature of the confidence scale being employed.

2.9.3.3 Ecological Validity

As discussed earlier, concerns have been raised about the ecological validity of findings from calibration studies. Very few have been undertaken in a classroom environment, with most being laboratory-based, in what could be argued as artificial situations (Hacker, Bol and Bahbahani, 2008; Kennedy, Lawton and Plumlee, 2002). Christensen-Szalanski and Bushyhead (1981) caution against generalising from experimental laboratory based findings to real life judgement scenarios, arguing that subjects in real life settings may not display the same cognitive limitations as those in a laboratory, as they may be more motivated to produce appropriate judgements in a real life setting. Hacker, Bol and Bahbahani (2008) therefore argue that, while experimental laboratory studies may be useful, more research should be undertaken in natural environments such as classrooms.

2.10 Reducing Overconfidence

In an analysis of a wide range of studies attempting to explain or reduce overconfidence, Fischhoff and MacGregor (1982) concluded that only three approaches appear to be effective: assigning easier tasks, challenging one's own judgements and extensive training supported by effective personal feedback. However subsequently, rewarding self-assessment accuracy (Russo and Schoemaker, 1992; Hacker, Bol and Bahbahani, 2008) and developing awareness of overconfidence (Russo and Schoemaker, 1992) have also been suggested.

2.10.1 Assign Easier Tasks

Since task difficulty has been associated with calibration accuracy, with difficult tasks tending to result in overconfidence and easy tasks underconfidence, asking subjects to perform less challenging tasks is likely to reduce overconfidence. However, this approach is of limited value since, while in experimental situations task difficulty could be manipulated, this is more difficult, as well as inappropriate in more natural settings in which individuals are faced with tasks of varying degrees of difficulty. In higher education for example, individuals may often be confronted with learning objectives which are designed to challenge them.

2.10.2 Challenge Judgements

Overconfidence can be reduced by considering evidence which contradicts initial beliefs (Zakay and Glicksohn, 1992; Hammond, Keeney and Raiffa, 1998) and actively considering reasons why judgements may be incorrect (Koku and Qureshi, 2004; Arkes et al, 1987; Koriat, Lichtenstein and Fischhoff, 1980), especially those made with 100% confidence (Dunning et al, 1990). This approach can be enhanced by the involvement of others (Russo and Schoemaker, 1992), since this provides another perspective on the judgement being made. The expectation that judgements may be scrutinised by others may also be influential. This can have a moderating motivational effect on those who may otherwise provide inappropriately high confidence levels for presentation purposes, due to their perception of the social utility associated with such displays. Evidence of

this was provided in a study by Arkes et al (1987), in which a group expecting a group discussion of answers they gave in a test displayed less overconfidence than a control group.

2.10.3 Develop Awareness

Fischhoff (1982) highlights the potential value of developing an awareness of overconfidence bias when suggesting the following progressive steps for correcting overconfidence bias.

1) Warn about the potential for the bias
2) Explain the direction and extent of bias which is typically found
3) Provide feedback on the subjects own bias to personalise the implications
4) Extensive training, coaching, feedback

He suggests moving through each of these in turn until the bias is corrected and Russo and Schoemaker (1992) explain that in some cases, awareness alone may be sufficient. For some this may result from general awareness of the tendency, prompted by Steps 1 and 2 above. However, for others this may be insufficient and Step 3, in which they are provided with information to make them aware of their own bias, may be more productive, while for others, the more extensive interventions in Step 4 may be necessary.

2.10.4 Feedback

Metaknowledge can be developed through the provision of relevant, timely and accurate feedback (Russo and Schoemaker, 1992; Lichtenstein and Fischhoff, 1980). Stone (2000) suggests that, although it may help to increase accuracy, the main value of feedback lies in its potential to induce individuals to adopt more appropriate confidence levels and it is this which enhances self-assessment accuracy. For example, it has been argued that evidence indicating that weather forecasters are particularly well calibrated,

with confidence levels matching very closely with accuracy (Charba and Klein, 1980), may be attributable to the rapid and highly relevant feedback they receive on their judgements (Murphy and Winkler, 1984). In separate studies of undergraduates on psychology courses, Nietfield, Cao and Osborne (2006) and Renner and Renner (2001) also found that providing feedback on performance in calibration tests reduced overconfidence and improved academic performance in classroom tests. However the type of feedback which may assist in moderating individuals' unrealistic evaluations of their own knowledge may not always be available to them. Kruger and Dunning (1999) argue that they may not receive any feedback and even if they do, it is more likely to be positive than negative, as those in a position to judge may prefer to provide none at all, rather than negative feedback. Therefore it is important that, rather than unduly positive feedback designed to enhance students' self-esteem, professional educators provide honest, constructive evaluations (Pintrich, 2002), which can assist learners in appreciating what they know and what they do not.

The effectiveness of feedback also depends on the extent to which recipients are able to use it effectively to inform future performance. Hacker et al (2000) for example, found that, while feedback improved the self-assessment accuracy of more competent students, this was not the case for the less competent, despite repeated information highlighting their limitations. Participants completed three tests of knowledge and both predictive and postdictive self-assessment accuracy over a 15 week period, in which they were informed of the potential benefits of accurate self-assessment and given feedback on test scores, as well as self–assessment accuracy. While for the group as a whole, predictive accuracy increased over time, postdictive did not. To investigate further they grouped participants, based on their competence in the knowledge based aspect of the test, with those achieving in excess of the median classified as high performers and those below as low performers. Their findings indicated that, while the higher performers improved over time in their predictive and especially, postdictive ability, the self-assessment accuracy of the less competent did not.

Keren (1987) argued that the value of feedback is influenced by the extent to which activities undertaken are related tasks, in which knowledge from one can be transferred to another. Weather forecasters' judgements of the likelihood of precipitation are examples of such tasks, since experience gained from making one judgement has

specific relevance for, and can be related to, subsequent judgements. Keren (1987) suggests that the process through which confidence ratings are generated in typical calibration studies consists of two sub-processes. In the first, the subject builds a mental model to generate feelings of certainty about, for example, the plausibility of one possible answer, versus the other options and these judgements are based on previous experience. In the second sub-process these feelings are quantified, by translating them into probabilities. He argues that where related tasks are performed repeatedly, their performance calls for similar cognitive processes. Consequently, subjects can learn from their previous experiences to become better calibrated. It is this which therefore contributes to very good calibration detected in studies of weather forecasters, who perform related tasks repeatedly. However, he suggests that the impact of feedback on the development of metaknowledge is likely to be more limited in respect of non-related tasks, because in their case, previous experiences are less informative. This argument calls into question the extent to which feedback may improve metaknowledge in a higher education context, as it would depend upon the degree to which the cognitive process associated with self-assessment of subject-related knowledge is transferrable between individual items of learning. It may, for example, be more easily achieved in environments in which learners focus on a particular narrow knowledge domain in one specialised subject area in which tasks are similar, than those which entail more diverse demands.

2.10.5 Training

The high quality of probability judgements provided by weather forecasters has been attributed, in part at least, to the fact that it is an important aspect of their role and great effort has been devoted to improving their judgements (Murphy and Winkler, 1984). However, while training can enhance metaknowledge (Lichtenstein and Fischhoff, 1980, Kruger and Dunning 1999), as discussed above, good calibration in one domain may not necessarily transfer to another. Keren (1987) for example, reported a study in which weather forecasters showed excellent calibration in their field of expertise, but performed no better than a group of students when participating in a general knowledge task. Lichtenstein and Fischhoff (1980) also reported little transferability of good calibration between general knowledge tests and other self-assessment tasks. Paris and

Winograd (1990) advocate the use of cognitive coaching in educational environments as a means of improving metacognitive skills generally and suggest that this could include self-assessment of learning as well as mutual discussion, instruction and encouragement to motivate the learners.

2.10.6 Reward Good Metaknowledge

Russo and Schoemaker (1992) argue that the lack of formal recognition or reward for metaknowledge, and failure to develop it during formal education, contributes to the general tendency for overconfidence. The fact that weather forecasters are rewarded for good calibration has been suggested as a factor contributing to findings indicating their highly developed metaknowledge (Fischhoff and MacGregor, 1982) and incentivising learners in educational environments, who display good metaknowledge, may assist in its development in that context also. In a study investigating this Hacker, Bol and Bahbahani (2008) awarded students sitting examinations on a course in educational psychology extra marks according to how accurate their predictions and postdictions of accuracy were. They found that, while accuracy improved for these students, this was also the case for a control group for whom there was no such incentive. Since there was no significant difference between the improvements in each group, the results indicated that the incentive had not been influential in enhancing calibration. They explored this in more depth, analysing the participants by competence by splitting them into two groups, based on whether their examination marks exceeded the median score. They found that for the higher performers, there was no difference between those offered the incentive of extra marks and those in the control group. However, for the lower performance group, the incentive did improve calibration accuracy, suggesting that explicit rewards may be more influential for such students.

2.11 Advantages of Overconfidence

While there is much emphasis in the literature on the possible negative consequences of excessive confidence, it is worth noting that potential advantages have also been highlighted. For example, as noted earlier, since confidence can be associated with

competence, especially in circumstances where competence is difficult to assess (Klayman et al, 1999), unwarranted displays of confidence can attract favourable attention and possibly rewards (Fischhoff, 1994). Overconfidence has also been linked to health benefits, with studies having found that self-enhancing biases such as overconfidence are more likely to improve psychological health than accurate self-perception (Beyer, 1990). This suggests that feedback to those with unjustifiably high perceptions of their own knowledge should be provided in a constructive and supportive manner. Confidence also has motivational value and encourages individuals to take on challenges, through which they may realise their potential. The German philosopher Goethe recognised this when declaring that, 'for a man to achieve all that is demanded of him he must regard himself as greater than he is' (Russo and Schoemaker, 1992 p.16). In an educational context, optimistic views of self competence can motivate individuals to apply themselves to learning and encourage them to undertake challenging learning activities (Paris and Winograd, 1990). This motivational impact can assist in improving performance and, as this further enhances positive self-perception, a positive cycle may be established in which self-evaluation, motivation and performance feed each other (Shen and Pedulla, 2000). Overly negative views of self-competence may lead to over-cautiousness and Davidoff (1995) for example, argues that in a medical context, an unwarranted lack of confidence contributes to excessive testing and at extreme levels, could result in the inability to take decisions. In education, it may diminish ambition and dissuade students from accepting challenges which may have increased their learning. Thus, since self-efficacy can engender the confidence necessary to take difficult decisions and rise to challenges which provide learning opportunities, it is important that it is not eroded in initiatives designed to heighten the awareness of learners who may be overconfident.

2.12 Conclusion

Summarising the issues discussed in this chapter, self-regulated learning highlights the importance of students taking responsibility for, and control of, their own learning and generally enhances performance. Metacognition is a concept which relates to 'knowledge and cognition about cognitive phenomena' (Flavell, 1979 p.906) and has been identified as the driving force behind self-regulated learning. While its components

are often inconsistently described, the most common distinction in the literature is that between metacognitive knowledge and skills, which have been described as knowledge about cognition and regulation of cognition respectively. One of the fundamental aims of education is to endow students with knowledge (Jehng, Johnson and Anderson, 1993) and an important aspect of metacognitive regulatory activity is knowledge monitoring, since accurate self-monitoring can enhance learning, particularly in environments such as higher education, which entail learning large amounts of information. However, learners are not necessarily well equipped to do this, a shortcoming which indicates poor knowledge of their own state of knowledge ('metaknowledge') and can impede their learning.

Calibration studies can be used to test for metaknowledge and the most popular approach is to use multiple-choice questions, in which respondents select, from a number of options, what they believe is the correct answer and provide a subjective probability representing their confidence in that choice. While many studies have been based on general knowledge questions, the ecological validity of this approach is questionable and the investigation of knowledge monitoring accuracy in more natural settings has been encouraged. This concept can be operationalised using the bias score, which provides information on both the extent and direction of the discrepancy between confidence and actual performance and is the most commonly used indicator of over or under confidence in knowledge. Previous studies have tended to support the view that metaknowledge is typically poorly developed, with a general tendency for overconfidence when individuals assess the extent of their knowledge. These findings have been attributed to the lack of formal recognition or reward for metaknowledge and failure to develop it during formal education (Russo and Schoemaker, 1992). However, exceptions have been reported with for example, good calibration being detected for meteorologists' weather forecasts, a finding which has been linked with training they undertake, the quality of feedback they receive on their judgements and the rewards available for accurate self-assessment. Task difficulty has also been found to be influential and high positive correlation between the difficulty of judgements and overconfidence has been widely detected.

Individual differences have also been reported, with previous studies indicating a greater tendency for overconfidence in males and younger people. A cultural bias has

also been detected, with Asian, and in particular Chinese, subjects displaying greater overconfidence than Westerners. In an educational context, the relationship between self-assessment accuracy and knowledge has also been explored, with many studies reporting that those who know more tend to have a better appreciation of their own knowledge than their less knowledgeable peers. However, this finding has been influenced by the fact that it has commonly been investigated by measuring both competence and bias using the same test instrument, which will tend to bias results towards a greater tendency for overconfidence in poor performers.

Given its possible adverse consequences, initiatives aiming to moderate overconfidence in knowledge may be useful and an understanding of its causes would be helpful when considering these. Reasons proposed include cognitive explanations and motivational factors, and suggested remedial strategies include raising awareness of the general tendency for overconfidence, challenging one's judgements, rewarding good metaknowledge and providing training and effective feedback.

2.12.1 The Current Study

The purpose of this study is to contribute towards answering central questions of how effectively individuals are able to self-assess their capabilities and what the implications of this are. It specifically focuses on their ability to appreciate the extent of their own knowledge. The preceding review of the literature has provided indications of what is known about this through previous studies, as well as highlighting issues which could be addressed through further research. This will now be used as a basis for the development of specific research questions and hypotheses for this study.

It has been suggested that an important issue, worthy of further investigation, is whether metacognition is a general concept, or specific to particular tasks and domains and that future research should therefore involve a variety of tasks and domains, as well as exploring individual differences (Veenman, Van Hout-Wolters and Afflerbach, 2006). This study addresses these issues by investigating a particular aspect of metacognition, knowledge monitoring, in a higher education setting.

2.12.1.1 The Tendency for Overconfidence

The most common finding in previous studies investigating the ability of individuals to assess their own knowledge accurately is the tendency for overconfidence. This is important in the context of learning, as the ability to appreciate what has already been learned and where gaps in knowledge exist may be advantageous in the development of learning strategies (Everson and Tobias, 1998). If students appreciate that their knowledge is deficient, they can undertake remedial action through further study, whereas ignorance of this is unlikely to induce such a response (Pintrich, 2002). While the literature review has indicated a general tendency for overconfidence in knowledge, its investigation in different settings has been suggested. In an educational context, the need for further research in a variety of institutions and academic disciplines has been highlighted (Kennedy, Lawton and Plumlee, 2002) and the lack of studies investigating overconfidence in business students specifically noted (Koku and Qureshi, 2004). The ecological validity of previous findings has also been challenged and it has been suggested that studies should focus more on natural settings, by investigating judgements made in domains with which subjects are familiar, such as knowledge acquired by students over a period of time in their natural learning environment. Grimes (2002) for example, calls for more research investigating students' ability to assess performance on tests in a classroom environment accurately.

This study addresses these issues by investigating knowledge monitoring accuracy of students studying in a higher education learning environment at a large business school, located in a post-1993 UK university. The investigation will focus on knowledge related to the study programmes on which participants are enrolled and specifically aims to answer the following question, by testing the associated hypothesis:

Research Question 1

Are students overconfident in their knowledge?

H_1 = *Students are overconfident in their knowledge*

2.12.1.2 Individual Differences

The literature review suggests that individual differences have been under-explored, in respect of metacognition generally and specifically relating to self-assessment of knowledge. Previous research on the tendency for inflated self-assessment has generally tended to focus on means, while not providing enough information on individual differences (Ackerman, Beier and Bowen, 2002) and Klayman et al (1999) suggest that individual differences in confidence judgements warrant further investigation. Grimes (2002) meanwhile, specifically advocates that this should address personal characteristics, such as age and gender and this study responds by investigating each of these. Additionally, since the research setting is a large business school educating students from diverse cultural backgrounds, differences by country of origin will also be explored.

Age

Views on how metacognitive skills develop over time are mixed. One theory suggests that they may emerge at an early age and increase steadily. Evidence for this has been found in studies indicating that self-appraisal of learning is enhanced and overconfidence in knowledge reduced, with age. However, evidence has also been reported, in which knowledge monitoring accuracy has reduced with age. The timing of the development of monitoring ability has also been considered, with one view arguing that it may develop later than other metacognitive skills, and another suggesting that it may be largely developed as a child and remain fixed during adulthood. Consequently this study aims to answer the following question:

Research Question 2

Is there an age difference in overconfidence bias?

H_2 = *Overconfidence differs between older and younger students*

Gender

There are consistent findings indicating greater overconfidence in knowledge for males. However, many studies have been based on general knowledge and it has been suggested that more research should focus on how monitoring accuracy may differ by gender in more natural settings, such as when answering learning tests (Lundeberg, Fox, and Punccohar, 1994). It has also been suggested that gender differences in educational settings may be influenced by the nature of the academic discipline under investigation. Consequently, the literature highlights the need to investigate these differences further and this study aims to do so by answering the following question:

Research Question 3

Are males more overconfident in their knowledge than females?

H_3 = *Overconfidence is greater for male students*

Country of Origin

Previous studies investigating differences by country of origin have detected particularly high levels of overconfidence for Asians, as compared with westerners (Acker and Duck, 2008; Yates, Lee and Bush, 1997). This has been particularly evident in Chinese subjects (Culpepper, Zhao and Lowery, 2002; Yates, Lee and Shinotsuka, 1996). This study aims to explore this issue by answering the following question:

Research Question 4

Are Chinese students more overconfident in their knowledge than UK students?

H_4 = *Overconfidence is greater for Chinese than UK students*

2.12.1.3 Overconfidence and Academic Performance

Grimes (2002) suggests that, since inaccurate self-assessment of knowledge may result in inappropriate learning strategies, and consequently, poor performance in assignments and examinations, the relationship between overconfidence and learning outcomes should be investigated. Nietfield, Cao and Osborne (2006) explain that previous studies investigating the association between calibration and performance have typically been restricted to performance in multiple choice objective tests. Koku and Qureshi's (2004) study of business students is an example of this, as they investigated the relationship between overconfidence and performance in multiple-choice examinations. However they used the same instrument to determine competence and overconfidence and as discussed previously, this approach is problematical. When determined in this manner, incompetence is more likely to be associated with overconfidence on the grounds that underconfidence is unlikely when the test accuracy score is particularly poor. This study aims to address this issue by using different datasets to determine indicators of self-monitoring accuracy and academic competence. While overconfidence bias is determined using multiple-choice tests, academic performance indicators are determined independently, using summative assessment marks achieved by students during their study programme. Consequently, the study aims to respond to the concern that previous studies investigating the association between calibration and performance have typically been restricted to multiple-choice examinations and little is known about how it relates to wider indicators of academic performance (Nietfield, Cao and Osborne, 2006). This will be achieved by answering the following question:

Research Question 5

Is there a negative association between overconfidence and academic performance?

H_5 = *There is a negative association between overconfidence and academic performance*

While their investigation focused only on a multiple choice examination, Koku and Qureshi (2004 p.223) claimed that their results 'could be generalized to all type of examinations'. However, they offer no evidence in support of this and make no reference to other modes of assessment. This study will address this by considering, not only participants' overall academic performance on their study programme, but also investigating the association between overconfidence and performance in different modes of assessment, such as assignments and examinations, as prompted by Grimes (2002). Since it includes higher education students at different levels of study, ranging from first year undergraduate to postgraduate, it also responds to Isaacson and Fujita's (2006) suggestion of exploring the relationship between metacognitive awareness and higher level learning.

Having reviewed relevant academic literature and developed specific research questions and hypotheses to guide the study, the next chapter will address the methodology and research design employed to provide answers to these questions.

3. Methodology and Research Design

3.1 Introduction

Research is concerned with asking and attempting to answer, questions in order to seek knowledge and develop an understanding of the world (Gill and Johnson, 1991). Social research investigates human activity and interactivity and includes educational research, which develops understanding of learning and teaching activities (Black, 1993). Various alternative approaches can be adopted in social research, each with relative strengths and weaknesses in terms of their assumptions, attitudes and underlying philosophies (Eilon, 1974). Consequently, it is important that researchers have an appreciation of the philosophical assumptions that influence how research is understood (Maykut and Morehouse, 1994). The manner in which it is conducted should also be clearly explained when reporting research and this chapter addresses these issues.

It commences by considering the role of theory in research before explaining the philosophy underpinning this study and the methodology adopted. This includes discussion of criticisms and support for the use of this methodological approach in social research, in order that the findings from the study may be interpreted accordingly. The research design is then explained in detail and mapped to the conceptual framework developed in Chapter 2. The nature of the research instrument used to collect data is explained, as well as piloting and data collection procedures. Issues potentially affecting the reliability and validity of the research are also considered, as well as initiatives implemented to address these, including the sampling strategy employed. Data analysis procedures are also described to clarify how data collected using the research design is to be interpreted and finally, ethical issues are addressed, to explain measures adopted to ensure that those involved in the study were not harmed as a result of their participation.

3.2 The Role of Theory in Research

Science is the means of understanding the natural and social world through observations and the process of science is usually classified as either inductive or deductive (Baker, 1999). An inductive process is one in which theory is constructed through the collection of data by repeatedly observing reality and developing explanations of what occurs (Saunders, Lewis and Thornhill, 2003). Thus, as illustrated in Figure 15, the process starts with observations, through which theory is generated using empirical generalisations. A deductive process, on the other hand, is conducted with the aim of testing theory. It therefore requires that a theory is identified, which is subsequently tested through empirical observation (Gill and Johnson, 1991). As shown in Figure 15, theories are used to generate hypotheses, which are tested through new observations (Baker, 1999).

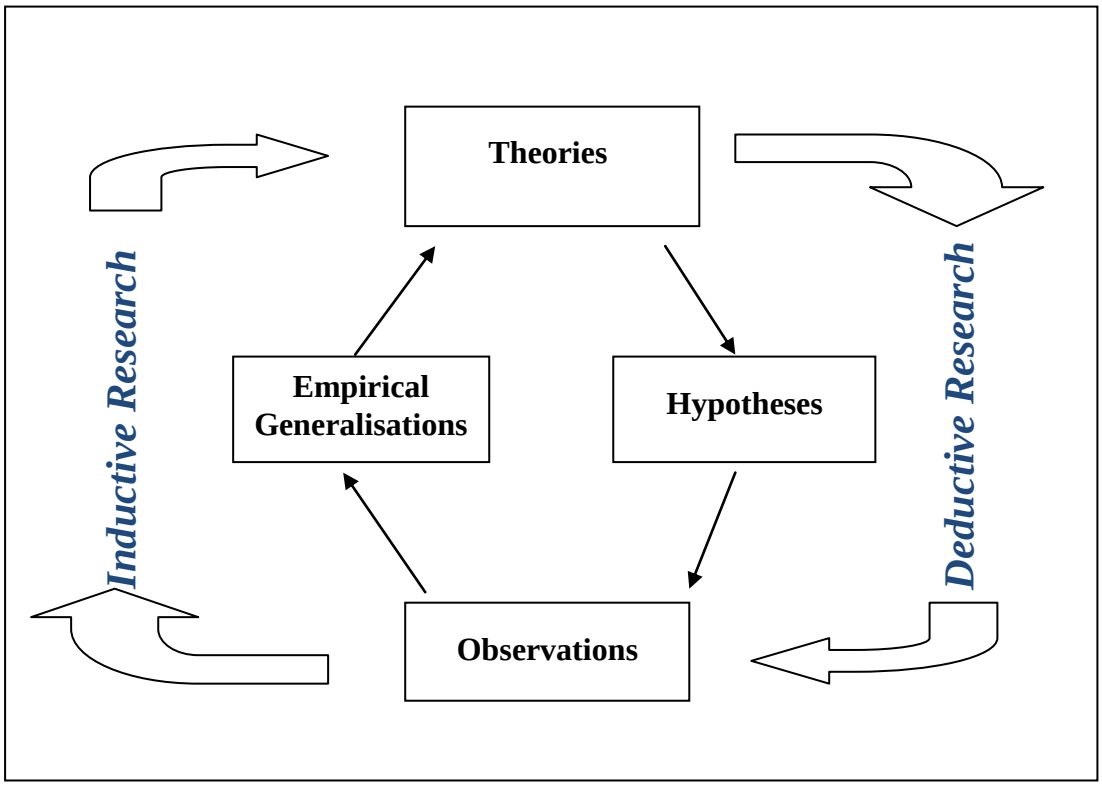


Figure 15: Components of the Scientific Process

Source: Adapted from Baker (1999)

This study involves a deductive process, as it tests theory related to self-monitoring of knowledge, which suggests a general tendency for overconfidence. As illustrated in Figure 16, this requires operationalising the concept of overconfidence by constructing an indicator to represent it (Hughes and Sharrock, 2007), which can be used to test theory by observing students studying business in a higher education environment. The methodology underpinning the study and research design used to carry out these procedures are addressed this chapter.

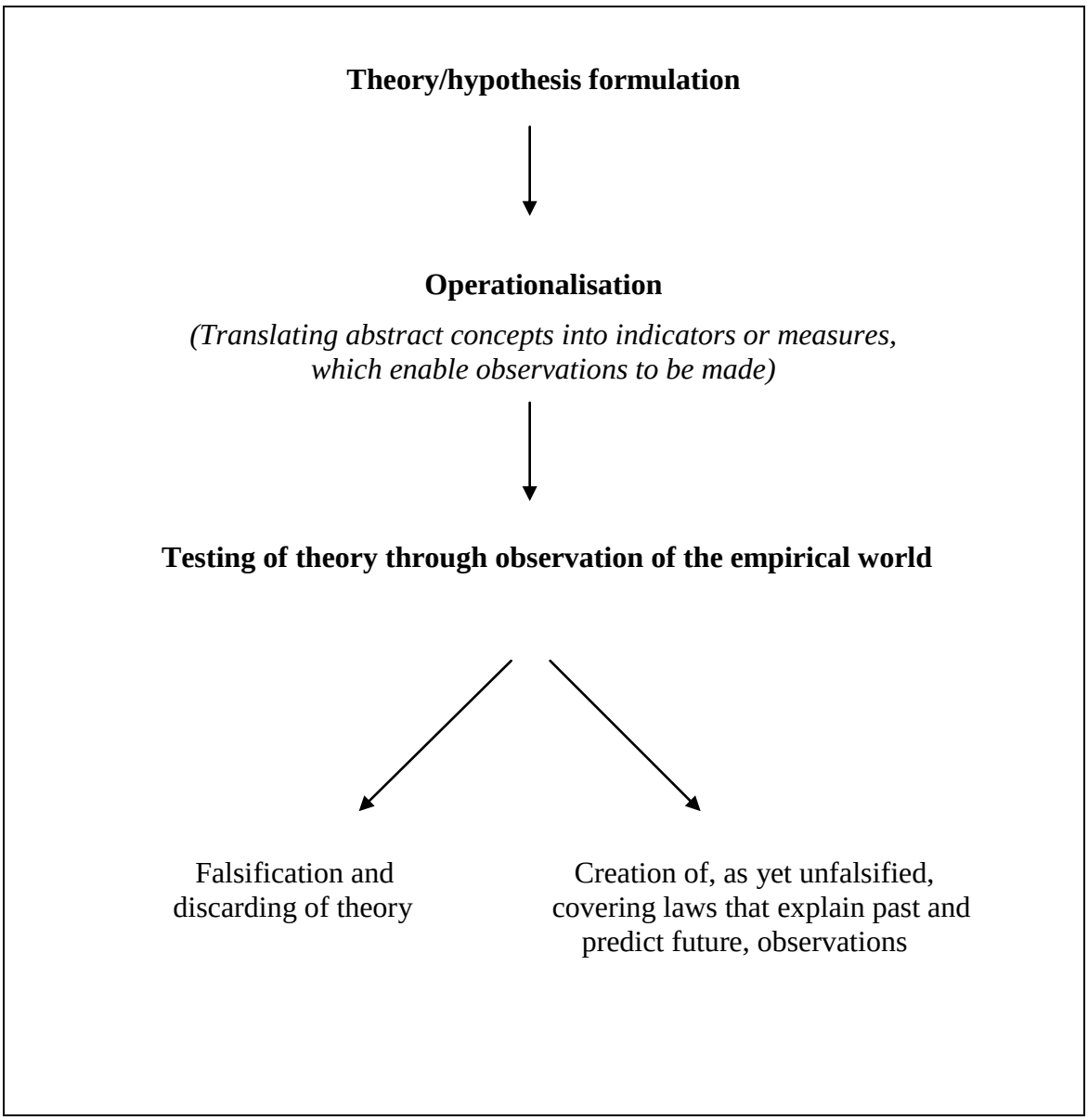


Figure 16: The Deductive Process

Source: Adapted from Gill and Johnson (1991)

3.3 Research Foundations

As researchers engage with and describe the world, philosophical issues underpin decisions made about methodological approaches (Scott, 2005). Consequently, when conducting research, it is important that the means used to derive findings are explicit and the assumptions under which they were produced made clear (Jankowicz, 1995). This should assist others in their interpretation, by providing an indication of potential biases in how data has been collected and interpreted as well as any other limitations in the work. However, while there are different approaches to research, the way in which these relate to theoretical issues is not always clear and often the literature uses terms in different, and occasionally contradictory, ways (Crotty, 1998). Sarantakos (2005) uses five elements to address how different approaches to research reflect a range of guiding philosophies and Figure 17 illustrates how the approach adopted in this study relates to these.

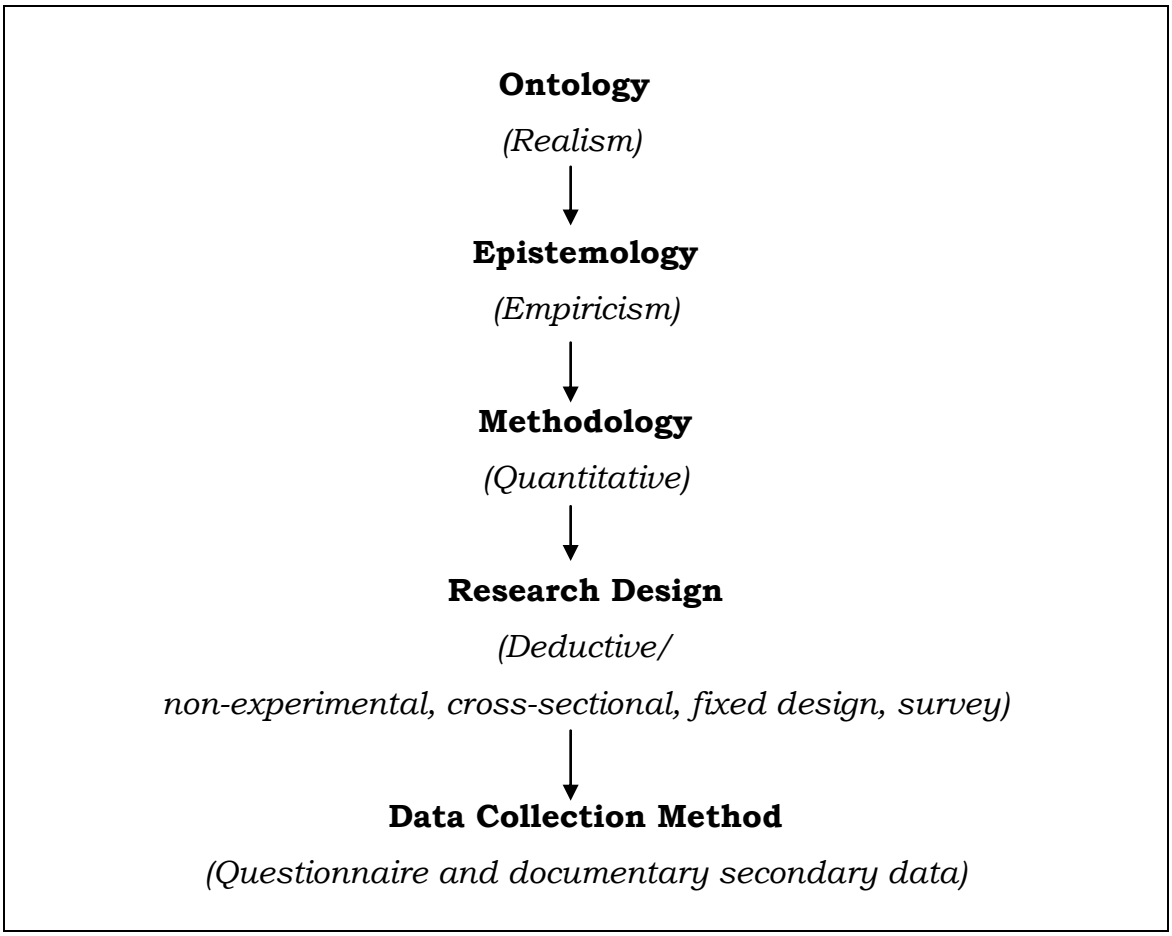


Figure 17: The Research Approach Adopted in the Study

3.3.1 Realist Ontology

The term ontology is derived from the Greek words ‘ontos’, which means being, and ‘logos’, which means theory or knowledge and thus, to consider the ontological status of something, is to question whether it is real or an illusion (Johnson and Duberley, 2000). It is therefore concerned with the nature of reality and addresses whether it is objective, or subjective and constructed. Realist ontology assumes that a reality exists, independent of human perception of it, whereas a subjectivist view sees reality as existing only through our conscious perception of it. Realism views reality as being governed by natural laws, knowledge of which would assist in predicting human behaviour (Sarantakos, 2005). This study is therefore informed by realist ontology in that it explores individuals’ metaknowledge, which exists regardless of whether those individuals, or others, are aware of it. Indeed by its very nature, inaccurate self-monitoring of knowledge, since it results from differences between perceived and actual knowledge, reflects a reality of which individuals are unaware.

3.3.2 Empiricist Epistemology

Epistemology is concerned with the criteria used to determine what constitutes warranted knowledge (Johnson and Duberley, 2000). Consequently, its contribution in activities in which knowledge claims are made is through ‘clarifying the conditions and limits of what is construed as justified knowledge’ (Johnson and Duberley, 2000 p.8). An empiricist epistemology deems knowledge to be created through facts, emanating from observed experiences, whereas interpretivism sees knowledge arising from subjective interpretations of the world (Sarantakos, 2005). Empiricism takes the view that knowledge arises through the senses, rather than other sources such as reason (Halfpenny, 1982), and is therefore created by gathering facts, through observation and experience (Sarantakos, 2005). Copleston (1963) highlights the relationship between realist ontology and an empiricist epistemology, when acknowledging the influence of experience in understanding the world by suggesting that, while truth precedes experience, it may be that we perceive it through experience. This study is informed by an empiricist epistemology in that the reality of overconfidence in knowledge and its

association with academic achievement are investigated in a sensory manner, rather than purely by reason, by observing evidence of each of these concepts.

3.3.3 Quantitative Methodology

Methodology holds a central position in research, since it is the strategy which translates ontology and epistemology into appropriate guidelines for its conduct (Sarantakos, 2005). In this respect, research is often classified as being either qualitative or quantitative. As illustrated in Table 3, a qualitative approach tends to accompany a constructionist ontology and interpretivist epistemology, which aims to interpret the different ways in which people make sense of their world. A quantitative approach, on the other hand, is typically guided by a realist, objectivist ontology, and an empiricist epistemology. It entails attempting to measure observations in empirical investigations and is typically underpinned by the positivist paradigm (Sarantakos, 2005).

Table 3 - Comparison of Quantitative and Qualitative Research Methodologies

	Methodology	
	Quantitative	**Qualitative**
Ontology	Realism/objectivism	Constructionism
Epistemology	Empiricism	Interpretivism
Research design	Fixed design	Fixed/flexible design

Source: Adapted from Sarantakos (2005)

3.3.3.1 Positivism

The term positivism is derived from the notion of a positive (i.e. progressive) approach and is originally attributed to the 19th century French philosopher August Comte. It was an influential intellectual trend from the mid 19th century and until relatively recently represented the commonly accepted view of science (Robson, 2002). Positivistic work incorporates statistical analysis and seeks causal relationships, or uses objectively

determined empirical information to test hypotheses (Halfpenny, 1982). However, as explained earlier, terms explaining approaches to research are used in different ways and while Sarantakos (2005) refers to it as a paradigm embracing realist, objectivist ontology and empiricist epistemology, Johnson and Duberley (2000) class positivism as an epistemological approach in itself and Hughes and Sharrock (2007) refer to it as ontology.

Positivism takes the view that it is possible to acquire accurate, value-free knowledge (Fisher 2004). It entails objectively studying and quantifying an 'observable reality', with a view to undertaking statistical analysis, to produce generalisable laws in the manner of natural scientists (Saunders, Lewis and Thornhill, 2003 p.83). It takes the view that knowledge is based on sensory experience and can be developed only by observation and experiments, using measurements to answer pre-determined questions (Cohen, Manion and Morrison, 2000). Thus, it reflects a structured, objective approach, typically using large samples and operationalises concepts for the purpose of measurement. Johnson and Duberley (2000 p.78) explain that this represents a nomothetic approach, applying procedures used in the natural sciences. They argue that, while positivism is the dominant approach in research in the management disciplines, this is often not apparent, since researchers rarely describe their work as positivistic. They explain that this does not imply that its assumptions are absent, but rather, that researchers may not explain their positivist rationale, because it is so dominant that they do not feel compelled to do so. Since methodology is closer to the practice of research, reference is more commonly made to this than the research paradigm, ontology or epistemology and therefore one may more commonly hear, for example, that a research study is quantitative, rather than positivistic (Sarantakos, 2005).

3.3.3.2 Criticisms of Quantitative Methodology

The use of a quantitative approach in the social sciences has been challenged and criticisms tend to reflect concerns with positivism (Sarantakos, 2005). A positivistic approach has been criticised in social research because it entails searching for constant relationships between variables and while this may be possible in the natural world, through the use of controlled experiments, Robson (2002) argues that the study of

people in the social sciences has yet to produce scientific laws despite attempts over more than one hundred years. Consequently, we cannot be sure that the results are truly representative in all instances and therefore represent certain knowledge. For example, the literature reviewed in the previous chapter identified a general tendency for overconfidence when individuals assess their own knowledge. However, there are exceptions, which indicate that this is not a consistent and universal finding.

Scott (2005 p.644) suggests that, as they operate with open systems, educational researchers have a more difficult task than those working in natural sciences, since individual behaviour and conditions in which it occurs, do not remain constant. He therefore argues that a quantitative approach to investigation of the social world is limited, since the need to express variables quantitatively results in descriptions that 'rarely reflect the richness and depth of human interaction'. Findings may be compromised by the manner in which research is designed. A quantitative approach is typically associated with fixed designs, in which the conduct of the research is pre-determined, and Sarantakos (2005) argues that this may over-restrictive which, along with a reliance on quantitative measures, can bias the findings of researchers. It can also present problems in terms of ecological validity, in that it may entail separating research participants from their natural context in order to study the issue in question, which may result in an oversimplification of the real world (Sarantakos, 2005).

3.3.3.3 Support for Quantitative Methodology

Some have argued that critics of positivist approaches have gone too far in abandoning scientific approaches and giving up hope of determining useful generalisations using quantitative approaches. Nash (2005 p.201) suggests that critics of a quantitative approach, risk 'throwing out the scientific baby with the positivist bath water' and cites Byrne's (1998) argument that a rejection of statistical methods is incompatible with the chaos and complexity which characterise the social world. Kemp and Holmwood (2003) support this view, arguing that events occurring in the social world are suitable for quantification, alongside non-positivist explanations. Cohen, Manion and Morrison (2000 p.27) also support the use of positivistic approaches, arguing that, while they 'can be criticised for their macro-sociological persuasion, so interpretive and qualitative

[approaches] can be criticised for their narrowly micro-sociological persuasion'. Denscombe (2003) meanwhile highlights the relative limitations of qualitative approaches when compared with quantitative methodology, pointing out that they tend to be less rigorous, over-descriptive, lack sufficient explanation and provide results that can be unrepresentative, due to the use of small sample sizes.

Johnson and Duberley (2000) explain that, despite its critics, many argue that the assumptions associated with positivism underpin most social science research and this view is supported by Sarantakos (2005), who argues that most social scientists still employ a quantitative methodology. However, Robson (2002 p.26) highlights the contribution of critics of positivism in recognising a 'looser connection between data and theory confirmation' than is assumed in positivism. A post-positivist view acknowledges their criticisms and accepts that, while positivism assumes that researchers are independent and objective, in practice their values and knowledge can affect research observations. It therefore seeks objectivity in research, while accepting that bias may impact on observations and interpretation and while believing a single reality exists, accepts that it can be known only probabilistically rather than perfectly, due to limitations of the research process (Robson, 2002). Fisher (2004) refers to this as realist research, which, while often deemed to be the same as positivism, can be usefully distinguished from it. While it retains many of the aims of positivism and aspires to a scientific approach, it accepts the subjectivity of research and therefore makes less strong claims for knowledge which perfectly reflects the objects being investigated. This post-positivist view therefore accepts that, while researchers may not be able to observe the world in completely objective manner, they should represent reality as best they can, acknowledging that findings will be influenced by subjectivity in conducting the research (Muijs, 2004). Thus while qualitative research may be informative in terms of why people behave as they do, quantitative studies can also play their part in understanding the social world. Nash (2005) challenges objections to the use of mathematical approaches in social research on the basis of intensionality, which suggest that a quantitative approach is inappropriate for uncovering meaning. He argues that social research should investigate events, regardless of whether the meaning of actors is known, and that objections to this would unnecessarily limit the scope of social science. Cohen et al (2000) support this view, arguing that, while understanding human

behaviour will be enhanced by an appreciation of intentions, this cannot be the sole purpose of social research. They suggest that, while patterns of social interactions may represent the accounts of actors, their consciousness may be false and cite Rex's (1974) argument that researchers have an obligation to discover an objective perspective, which may not accord with that of the actors. The very focus of this study addresses a theory that humans may overestimate their knowledge, one which, in itself, suggests the limited consciousness to which Cohen et al (2000) refer. Consequently, here, it is investigated using quantitative methodology, and while the study attempts to achieve the objective perspective to which Rex (1974) refers, the difficulty in doing so, due to the limitations of the research process, is acknowledged.

Pring (2000 p.259) argues that both quantitative and qualitative research have a part to play in educational research. A quantitative approach can be used to suggest differences, which can be explored in a more interpretive manner, to refine general claims and determine the particular meanings behind actions and the beliefs which underpin the 'quantifiable claims which research should constantly be seeking'. While acknowledging the role of qualitative methodologies, he concludes that only by employing quantitative approaches, using statistical methods, will educational practitioners have access to (perhaps sometimes tentative) conclusions, to assist them in understanding which interventions are likely to be useful. His view is therefore supportive of the approach taken in this study, which aims to inform professional practice, by investigating the extent of overconfidence in knowledge in a higher education setting and its association with academic performance. Testing existing overconfidence theory in this manner responds to Kennedy, Lawton and Plumlee's (2002) call for further research investigating it in different institutions and disciplines, and will help to assess the generalisability of previous findings and inform practitioners, who have professional responsibility for enhancing the learning of business students. However, while adopting a quantitative methodology, it is acknowledged that the study may not comply with the more strict assumptions associated with positivism concerning complete objectivity, researcher independence and the ability to operationalise concepts unambiguously. Consequently, to assist in the interpretation of the research findings, it is important that the manner in which the research has been conducted is clarified and this is explained below.

3.4 Research Design

Research design is a process through which research is structured to answer research questions (Davis, 2005) and different types of research lend themselves to different means of data gathering (Jankowicz, 1995). Research designs may be fixed or flexible and quantitative studies typically use fixed designs, in which the approach to the research is largely determined in advance. These are theory driven and tend to rely on statistical interpretation of quantitative data. They usually consider aggregates, tendencies and group properties and their advantage is in their ability to identify patterns and processes, which can be associated with social structures and features of groups (Robson, 2002). In support of the research philosophy adopted in this study, a fixed design was employed to provide indicators of metaknowledge and enable the investigation of patterns, by exploring individual differences and relating overconfidence bias to indicators of academic performance.

When using fixed designs, a distinction can be made between experimental and non-experimental (or 'correlational') strategies. These are similar in that they are empirical, with evidence gathered by observation and measurement, and attempt to answer research questions objectively (Field and Hole, 2003). However, they differ in that non-experimental methods attempt to gather data from the real world without attempting to interfere with it directly, whereas an experimental approach involves manipulating the environment in some way (Field and Hole, 2003). An experimental approach entails a systematic approach to research, in which subjects are allocated to experimental and control groups, and the researcher manipulates independent variables and observes changes in dependent variables (Gill and Johnson, 1991). However, in social research, such experiments can be difficult to conduct outside of the laboratory. The use of an experimental approach in social science also presents potential ethical issues, where for example, groups are treated differently for control purposes. Koku and Qureshi (2004) for instance, randomly assigned students taking an examination into one of two experimental groups or a third control group. The two experimental groups were respectively requested to provide possible reasons contradicting, or supporting, the answers they gave in order to test hypotheses that the former would reduce overconfidence and the latter increase it. The results from the study demonstrated that

for the ‘contradicting reasons’ experimental group the intervention did indeed result in lower overconfidence than each of the other two groups. That group also achieved a mean test score of 72%, which exceeded that of both the ‘supporting reasons’ experimental group (62%) and the control group (65%). If these differences in test scores were attributable to the experimental interventions, then it could be argued that the ‘supporting reasons’ group were harmed by the research, since they achieved a lower mean mark than each of the other two groups. This study does not entail manipulation of variables or assigning participants to experimental and control groups and therefore, since it uses a non-experimental design, avoids such ethical issues.

Robson (2002) distinguishes between three different types of non-experimental, fixed designs. In comparative designs, participants are grouped and the emphasis of the study is on making comparisons between those groups. Longitudinal designs, on the other hand, study the same subjects and measures are repeatedly taken over a period of time in order to analyse changes and trends. The third type is relational designs, also known as correlational studies, which entail taking measurements on a range of variables and examining the relationship between them. A common method for doing so, which is often used in conjunction with a survey, is a cross-sectional design, in which measures are taken across a relatively short period of time. This approach, which studies the relationships between and among variables in a group of subjects, is probably the most commonly used design in social research (Robson, 2002) and was adopted in this study. Due to the problems associated with determining causality in relational designs, the terms, independent and dependent variables are often replaced by terms ‘explanatory’ and ‘outcome’ variables and while the detection of causal links is difficult, the relationship between variables may in itself be of interest (Robson, 2002).

In this study, a fixed design was used to investigate associations between an outcome variable indicating overconfidence bias and those indicating academic performance. While this entails relational analysis, the aims of the study also included investigation of explanatory variables representing individual differences in overconfidence in respect of age, gender and country of origin. It therefore employs a cross sectional design which is essentially a combination of a relational and comparative study within the same framework.

3.4.1 Evaluating Overconfidence

A deductive process entails operationalising concepts and testing theory through empirical observation. Concepts are labels attached to 'elements of the social world that seem to have common features and strike us as significant....and represent the points around which social research is conducted' (Bryman, 2008 p.143). Measuring them provides a basis for making distinctions and allows investigation of differences between people, in terms of the concept being studied and relationships between it and other concepts (Bryman, 2008). This study investigates the concept of metaknowledge, through observations of self-monitoring accuracy in a calibration study. It is therefore necessary to operationalise metaknowledge to determine its extent in students and permit the investigation of differences between various sub-groups, as well as the association between this concept and academic performance. When doing so, it is important to distinguish between measures and indicators. The former relate to things which can be clearly quantified, such as age, or number of years worked in an organisation, whereas indicators are used where concepts are less directly quantifiable. They allow more abstract concepts to be quantified and subsequently analysed as if they were direct measures (Bryman, 2008). Thus, it is more appropriate to refer to the variable used to operationalise metaknowledge as an indicator, rather than a measure, as it is often referred to in the literature (e.g. Pieschl, 2009; Ehrlinger et al, 2008).

As discussed earlier, there are two types of calibration study, each of which establishes a different indicator for metaknowledge. The first uses range questions in response to which subjects provide upper and lower estimates of numerical values, which correspond to their belief that the correct answer has an x% chance of lying between those values. Since this approach entails using questions requiring a numerical response, using it in this study would restrict the breadth of knowledge that could be addressed. However, the second approach overcomes this problem since it requires subjects to make judgements about discrete propositions and provide confidence estimates in respect of each judgement. This is less constraining in terms of the nature of the knowledge which can be addressed and since the aim of the study is to investigate metaknowledge in the context of issues addressed in the participants' study programme, this approach was adopted.

Schraw (2009) recommends that the outcome indicator used when assessing metacognition should be that which relates most closely to the aims of the research. Consequently, as the research questions in this study relate to levels of over/underconfidence, the bias score was employed as expressed below:

$$\text{Bias score} = \frac{1}{N} \sum_{i=1}^{N} (c_i - p_i)$$

Where:

N = the total number of responses

c_i = a confidence rating

p_i = a performance score (rated as either 100% (correct) or 0% (incorrect)).

As discussed earlier, this is the most commonly used indicator when assessing overconfidence (Ehrlinger et al, 2008; Hacker et al, 2000). It represents 'the difference between the mean of the probability responses and the overall proportion correct' (Lichtenstein and Fischhoff 1977, p.161) and thus, entails a comparison of the mean confidence expressed across all judgements and the success rate. The latter is referred to as the performance score in the formula above and for the purposes of this study, to avoid confusion with aspects of academic performance, will be referred to as the 'knowledge score'. Thus the derivation of the bias score can be simplified as follows:

$$\text{Bias score} = \text{MC} - \text{KS}$$

Where:

MC = Mean confidence across all judgements

KS = Overall proportion correct

The bias score indicates both the direction and size of judgemental error. A positive score indicates overconfidence and a negative result indicates underconfidence and its size indicates the severity of the judgemental error (Schraw, 2009). It is also worth noting that when providing self-assessment feedback, Farrell and Leung (2004) stress the importance of engaging the users by adopting an approach which is not cognitively over-demanding and adopts a relatively straightforward scoring system. The bias score complies with this and while this in itself does not impact on the conduct of this study, it does have potential implications for its use in initiatives in professional practice, which may be suggested by findings generated by the research. Dissemination of feedback using a relatively simple indicator is likely to more acceptable not only to students but also to professional practitioners in higher education.

3.4.2 The Research Instrument

Research instruments are the means used to collect data for the purpose of analysis (Sarantakos, 2005) and in this study a questionnaire was used (see Appendix 8). While these provide a quick, inexpensive means of collecting meaningful data (Bell 1993), response rates are often low (Wisker, 2001) and many produce response rates below 50%, with virtually none achieving 100% (Muijs, 2004). To help overcome this, questionnaires should be attractively designed (Bell, 1993) and concise, while allowing the collection of sufficient relevant data to meet the objectives of the research (Gill and Johnson, 1991). The questionnaire used in this study was therefore clearly set out and contained a set of instructions to guide the participants, as well as a clear indication of where responses should be recorded, which would also simplify the data recording process. To facilitate subsequent statistical analysis, closed questions were used, as recommended by Jankowicz (1995). The questionnaire comprised two sections in order to collect data which could be used to determine overconfidence, as well as demographic data to permit the investigation of individual differences. The design of each of these two sections is discussed below.

3.4.2.1 Demographic Data

As suggested by Gill and Johnson (1991), the demographic factual data was requested first. Respondents were asked to indicate their age group, gender and country of origin,

to allow investigation differences according to these characteristics. This information would also permit subsequent checking of the representativeness of the sample in these respects (Fisher, 2004). They were also asked to provide their unique student reference number and indicate their study programme. The student number permitted access to the secondary data required to determine their academic performance on their study programme. Using their reference numbers, rather than names, for this purpose helped to protect their anonymity and as well as complying with ethical standards for research, this was designed to improve the response rate of those willing to allow their data to be used for the study. The reference number also enabled the determination of the number of years they had spent studying at the university, since the first two characters are significant in this respect. Participants were also asked to record the name of their study programme, as this could be used to assist in determining the relevant student registration number should a participant not indicate this clearly enough to correctly interpret.

3.4.2.2 Testing for Overconfidence

The second part of the questionnaire comprised the instrument used to determine the bias score and contained 30 multiple-choice objective test questions. As is typically the case for such questions, each comprised a stem, indicating the problem and a number of alternative responses, including one which was correct and a number of distracters (Hansen and Dexter, 1997). The position in which the correct answer appeared across the 30 questions was randomised using a random number generator. Using questions for which there is a definitive correct answer in this way overcomes the difficulty of the lack of an objective benchmark when investigating self-assessment (Kruger and Dunning, 1999). Respondents were requested to select the correct answer for each question and indicate their confidence in that choice.

The written instructions provided them with guidance in terms of how the confidence scale was to be used, with 100% indicative of their certainty that the answer provided was correct and 25% to be used in cases where they felt they were only able to guess and therefore had a one in four chance of being correct. These instructions also provided an explanation of how 50% confidence should be used in situations where the respondent was certain that two options were incorrect and believed that the other two

had an equal chance of being correct. An additional illustrative example showed a confidence level of 60% and it was explained in verbal instructions provided to participants by the researcher prior to completing the test that this could for example, be used in circumstances in which they believed that the correct answer was one of two possible options of which they judged that one was more likely than the other. Consequently, while the written instructions provided illustrative examples to guide respondents on the use of the scale, the verbal explanation by the researcher permitted other scenarios to be addressed. These included for example how 90% confidence could be used where respondents had a strong belief that they had chosen the correct answer, but acknowledged that there was a possibility (i.e. one in ten in this case) that this was not the case.

The approach adopted is mapped onto the conceptual framework developed in Chapter 2 in Figure 18. In terms of granularity and timing, as participants were required to provide confidence estimates for a series of test items after having completed each question in turn, this study employs the most popular approach used in calibration studies (Keren, 1991), by investigating local retrospective judgements, using an online testing approach. As well as a bias score indicating the extent of overconfidence for each respondent, this approach also permitted the production of calibration curves to illustrate monitoring accuracy for the whole sample and various sub-groupings at different levels of confidence.

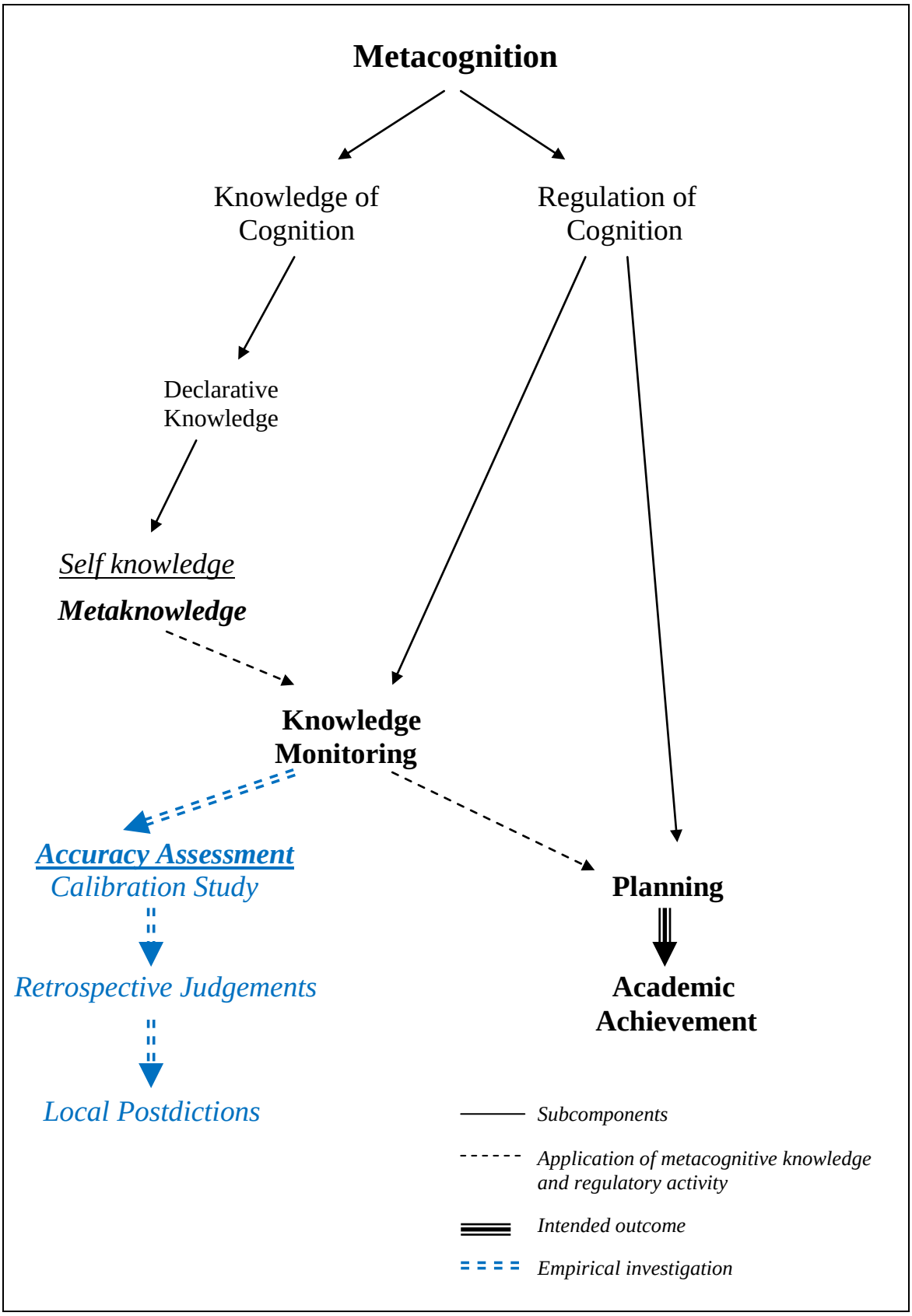


Figure 18: Mapping of the Empirical Investigation onto the Conceptual Framework

3.4.2.3 Knowledge Domain of the Test

As explained earlier, the ecological validity of previous studies has been challenged on the basis of their emphasis on general, or experimentally learned, knowledge rather than richer knowledge gained in a more natural learning environment (Hacker et al, 2000). Consequently, the knowledge base used to test for metaknowledge in this study was specific to the study programme on which participants were enrolled and therefore, as suggested by Dunning et al (1990), enabled the investigation of judgements in domains with which the participants were familiar. To permit the investigation of students studying at different levels, eight separate tests were designed covering four levels of study, ranging from first year undergraduate up to postgraduate level. Since the students in question study various business related disciplines and subjects (marketing, finance, human resource management etc.), in order to control for a possible subject-related confounding effect, each test focused on a study module which addressed financial and quantitative aspects of their study programme.

3.4.2.4 Question Difficulty

Since the use of deliberately misleading test items can create an experimental bias, which tends to produce overconfidence (Keren, 1991), questions were designed to provide a test of knowledge which participants could have reasonably been expected to have attained. To facilitate this, they were compiled by academic staff at the university with relevant subject expertise and responsibility for the study modules on which the tests were based. However, this did not guarantee tests of equal difficulty and it was therefore necessary to investigate differences between them in this respect statistically, when analysing the data collected. This would permit consideration of steps to control for a possible task difficulty confounding effect when considering individual differences in overconfidence.

3.4.2.5 Number of Distracters

While the use of two-choice questions, in which respondents choose from one of two possible answers, has been the most common approach in confidence studies (Klayman

et al, 1999), Keren (1991) suggests increasing this number, to represent more fully different knowledge states of the respondents. However, he cautions against too many alternatives on the grounds that limited cognitive processing capacity would prevent respondents from making meaningful comparisons between them. Consequently, in this study, a compromise was reached by using four-choice questions, which would increase the scope for respondents to transmit information (Shuford and Brown, 1975) as compared with two alternative items.

3.4.2.6 Eliciting Confidence Levels

As explained above, confidence judgements were elicited using numerical expressions rather than verbal responses. As forced choice judgements using multiple-choice questions require the use of confidence estimates between 1/n and 1 and in this case four alternatives were provided, participants were instructed to assign confidence levels in the range 25% to 100%. A response of 100% represented certainty in cases where they felt sure they knew the correct answer and 25% was to be used where their lack of knowledge meant they could only guess and the likelihood of doing so accurately was therefore down to chance. While previous studies have often restricted confidence responses to 5% intervals (100%, 95%, 90% etc), Fischhoff, Slovic and Lichtenstein (1977) advocate the use of a more graduated scale and therefore in this study, respondents were instructed to use any number between 25% and 100%.

It has been suggested that overconfidence reported in previous studies may result from respondents not understanding the task and in particular, the use of the probability scale through which confidence levels are communicated (Keren, 1991). Consequently, studies such as this may be compromised to some extent by respondents' difficulty in translating subjective probabilities into numerical representations. Response contraction bias, which was discussed earlier, is an example of this, as is Ronis and Yates' (1987) finding that some respondents provided probability estimates at below chance level. In the latter example, the researchers addressed the issue by manipulating responses in a manner which they believed better reflected the judgements of participants. However, Keren (1991) suggested that such respondents may not understand the probability scale, in which case they should be removed from the sample. Consequently, in this study,

respondents who, through their use of confidence levels of less than 25%, appeared not to understand the use of the probability scale, were excluded from the sample.

3.4.2.7 Number of Questions

Where the intention is to study each respondent separately, a large number of observations are required in order to enhance the reliability of the results. Increasing the number of questions used also reduces the potential for the test to be biased towards contrary questions, which is more likely to result in overconfidence. Where respondents do not all complete the same test, as is the case in this study, increasing the number of questions also permits using a smaller sub-sample of questions from each test to derive tests of approximately equal difficulty. This procedure will be considered further later in this chapter, when discussing approaches to data analysis. However, increasing the number of questions also has its problems, such as the time required to complete them and the potential for respondents to become bored while doing so (Keren, 1991) and therefore not engage appropriately with the process. Since in this study, the test was conducted during the participants' scheduled study programme, it was designed to be completed within a single one hour session. Approximately 25 minutes were allowed for clarifying the nature of the task with respondents at the start of the session, allowing 35 minutes for recording demographic data and completion of the test questions. Consequently, on the basis of the demographic section of the questionnaire requiring no more than five minutes and each test question requiring approximately one minute to complete, 30 questions were used in each test to provide sufficient time for completion within one hour.

3.4.2.8 Piloting the Research Instrument

When using fixed designs, it is important to carefully design the methods to be employed to answer the research questions before commencing data collection. In this study, a questionnaire was used and these have the advantage that they can permit the collection of data relatively quickly and consistently, particularly where closed questions are used. However, they can be difficult to design well (Wisker, 2001) and piloting them prior to their distribution can help to identify any problems respondents

may have in completing them (Jankowicz, 1995). Consequently, the research instrument to be used in the study was piloted with a group of 40 postgraduate students. These mainly comprised non-UK students, whose first language was not English, as it was anticipated that despite the necessity to satisfy the university as to the standard of their English language skills, such participants may have the greatest difficulty in interpreting the task requirements. The piloting activity was conducted in a one hour session and was followed by another session the following week, in which completion of the questionnaire was discussed with participants. This confirmed that one hour was sufficient time for an introductory explanation of the task and for respondents to complete it, since they all submitted their questionnaire within that time period. Analysis of the data also indicated that none had used a confidence rating below 25% and thus, all appeared to appreciate the use of the scale in this respect. However, discussion of the task with the group did highlight the potential for some confusion when providing confidence levels. For illustrative purposes, a hypothetical situation was discussed in which, for a particular question, one of the four alternative answers available is judged to be incorrect and the remaining three are deemed to have an equal chance of being correct. While the majority of the group appreciated that this corresponded with a 33% (one chance in three) confidence level, one person believed they had a 75% chance of being correct in this situation. Consequently, this was taken into account in the explanation of the task at subsequent data collection events, where more emphasis was placed on the use of the confidence scale and the implications of this hypothetical situation, along with others was discussed. These were illustrated using practical demonstrations which entailed retrieving one of four different coloured pens from a container, in which the participants were asked to make confidence judgements in respect of different scenarios. Given the feedback from the piloting activity, these included one for example, in which, in full view of the participants, one pen was removed from the container. They were then asked to indicate their confidence level in choosing a given colour from the remaining three. These demonstrations were interactive and supported by inter-participant discussion as well as group discussion involving the researcher.

Another initiative was introduced at this stage in respect of the confidence judgements, which was designed to address the potential problem of response contraction bias. As

discussed earlier, it has been argued that this can arise where respondents who are particularly uncertain about the answer to a test item provide a response around the mid-point of the scale available to them (around 60-65% in the case of this study), rather than indicating a confidence judgement nearer chance level. To address this, an additional diagnostic question was incorporated in the test as a means of identifying those who appeared not to be using the confidence scale appropriately. As shown in Figure 19, this question was one for which it was highly unlikely that respondents would know the correct answer and to reinforce this, they were provided with four options with values which were very closely grouped. Thus, it was designed in such a manner as to prompt a guess, which should therefore have been accompanied by a confidence level of 25%.

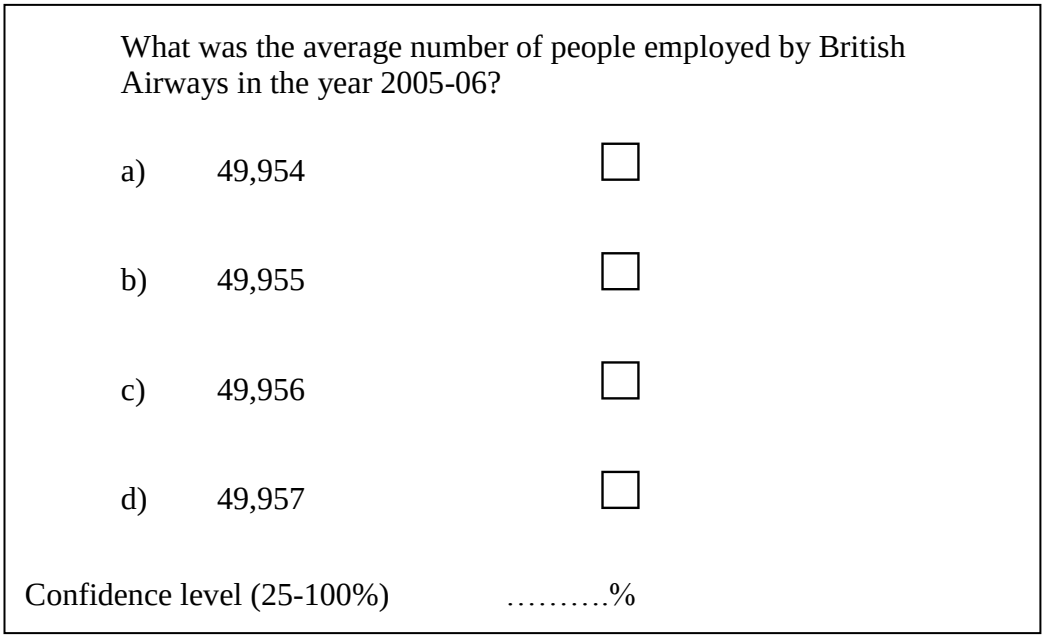


Figure 19: Diagnostic Test Item to Indicate Inappropriate Use of the Confidence Scale

When discussing the introduction of this question with the pilot group, while the majority indicated that a confidence level of 25% was appropriate, one indicated 50% confidence. When questioned about this, they explained that they thought they could make a more informed judgement than merely relying on chance, due to their (incorrect) belief that the average number of employees in the year must be an even number. Consequently, this question was amended, as shown in Figure 20, to one which respondents were less likely to interpret as being answerable using informed judgement.

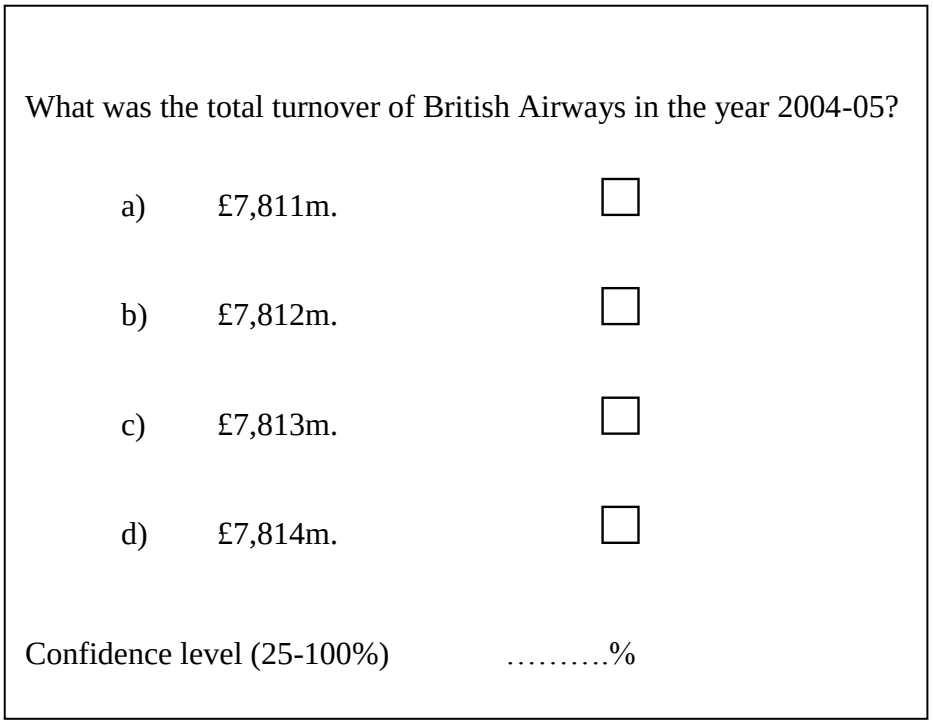


Figure 20: Adapted Diagnostic Test Item to Indicate Inappropriate Use of the Confidence Scale

Any respondent providing a response other than 25% for this question was eliminated from the study, on the grounds that it appeared that they were using the response scale inappropriately (due to response contraction bias or any other explanation). This question was employed solely for this diagnostic purpose and was excluded from the data used to determine bias scores. While these initiatives may not necessarily have completely eliminated problems associated with the use of the confidence scale, they were designed to assist in doing so and therefore, in enhancing the reliability of the data collected.

3.4.3 Data Collection Procedure

3.4.3.1 Primary Data

Primary data is 'collected for a specific purpose from original sources' (Davis, 2005 p.270). A potential problem in collecting it using questionnaires is a low response rate.

A lack of participants in calibration studies may mean subjects being evaluated as a group, rather than on an individual basis (Tomassini et al, 1982) and as this study aims to study individual differences, a good response rate was desirable. To improve response rates from questionnaires, it is preferable to distribute them personally (Bell, 1993). Loftus and Wagenaar (1988) reported the effectiveness of this approach, as well as the benefits of accessing potential respondents at a time when it is easier to hold their attention, when discussing a study investigating overconfidence among lawyers attending a series of educational seminars. Initially they left questionnaires at the registration desk, where a facilitator invited the target group to participate, but this approach produced a response rate of only 5%. However, a modified strategy, in which the target group was supplied with the questionnaire and asked to participate once they had sat down in the seminar room, produced a response rate in excess of 80%. Therefore, in this study, data was collected at a session which respondents attended during the course of their study programme. At each session the questionnaires were personally distributed by the researcher and, given the relatively short time required for their completion, collected personally shortly after they had done so. Each session was also personally supervised by the researcher, to ensure that participants did not collude. Since it has been suggested that overconfidence detected in previous studies may have resulted from respondents failing to fully understanding the nature of the task and in particular, the use of the confidence scale, the questionnaire provided specific guidance in respect of the confidence judgements requested (See Appendix 8). However, the complexity of information required from respondents in questionnaires may also require the presence of the researcher to explain the requirements (Gill and Johnson, 1991). Consequently, before each test the researcher clarified the task requirements, placing particular emphasis on the requirement for confidence judgements. As explained earlier, this included an interactive discussion regarding the use of confidence levels to reflect uncertain outcomes. It was also emphasised that students were under no obligation to consent to their data being used in the study and their decision as to whether they permitted this would have no bearing on the summative assessment of their performance on their study programme or any other adverse consequences.

3.4.3.2 Secondary Data

Secondary data is 'data that has been collected by others for another purpose' (Davis, 2005 p.70). As this study aimed to investigate the association between academic performance and overconfidence bias, it was necessary to obtain secondary data indicating the academic performance of participants from the university's marks recording system. This would permit comparison of overconfidence with an independently derived indicator of competence. As discussed earlier, it has been suggested that the tendency for the less competent to also demonstrate poorer self-monitoring accuracy in previous studies has been attributed, at least partly, to the use of the same instrument to assess both metaknowledge and competence. This arises when competence is determined by using the correct response rate to the questions in the test. In these conditions it is more difficult for those demonstrating poor knowledge, as indicated by a low test accuracy score, to underestimate this knowledge and they are therefore more likely to display overconfidence than those displaying better knowledge. Consequently, in this study, this was addressed by using independently derived indicators of competence based on participants' academic performance. These were determined using marks achieved during the stage of their study programme in which the test was conducted and for each student participating in the study, a percentage mark was obtained which reflected their overall academic performance in that stage. Koku and Qureshi (2004) argued against the use of an overall score, on the grounds that it is a composite performance measure and evaluation of students occurs on different bases, depending on the nature of their own specific study programme. While this is also the case for this study, its impact is mitigated by the fact that the institution uses general assessment criteria, which promote consistency in the assessment process at each level of study, regardless of the mode of assessment. As well as the indicator of overall performance for each student, further data was collected which reflected component marks obtained for each of three different modes of assessment; examinations, oral presentations and coursework in which students are given assignments to work on over a specified period of time. In the case of final year undergraduates and postgraduate students, a further sub-component of the coursework category, which related to their mark on a research dissertation, was also obtained. This indicator was not relevant for lower level undergraduate students who are not required to produce a dissertation.

Obtaining this additional data permitted investigation of the association between overconfidence bias and performance in each of these types of assessment, as well as overall academic performance.

3.4.4 Replicability

Bryman (2008) argues that three of the most important criteria for evaluating social research are replicability, reliability and validity. Replicability relates to the extent to which other researchers can replicate the study, perhaps in a different time period, or in a different setting. It requires that studies are conducted in a manner which permits other researchers to repeat them in an objective manner, without being subjectively influenced by the researcher, in order that comparisons can be made between studies in order to generate legitimate generalisations (Sarantakos, 2005). For example, this study is informed by methods adopted in previous research discussed in Chapter 2, and operationalises overconfidence using the most commonly used indicator, the bias score (Ehrlinger et al, 2008; Hacker et al, 2000). The explanation of the conduct of the study in this chapter and the inclusion of the research instrument employed assist in its replicability in that other researchers who may wish to adopt this approach in other settings, can be guided by this information.

3.4.5 Reliability

While it is important to address them, threats to reliability and validity cannot be entirely overcome and therefore, researchers should rather focus on how they can be reduced (Cohen, Manion and Morrison, 2000) and acknowledge them when reporting on the research. Reliability is concerned with how consistently a concept is measured and thus whether a measure of a concept is stable (Bryman, 2008). Reliable research is such that, if conducted in a similar context with similar participants, it should produce similar results (Cohen, Manion and Morrison, 2000). However, it is not easily established and it is difficult to solve potential problems associated with attempts to do so (Bryman, 2008). One approach to testing reliability in this study would be to collect data from the same participants more than once and investigate whether the correlation

between the results is high, thus indicating stability. However, Bryman (2008) explains that a problem in this approach is that the experience of undertaking the test the first time may influence the responses provided in the second test. Another problem relates to the influence of the time elapsing between the two tests. Inconsistent results between each test may be due to the influence of experiences occurring in the time period between them and not the research design and because of these difficulties, most research studies do not incorporate tests of stability (Bryman 2008). However, initiatives can be implemented which attempt to prevent causes of unreliability when designing the research and these are considered below for this study in respect of three important factors, participant bias, observer bias and situational factors.

3.4.5.1 Participant Bias

Since participants' declarations of their confidence for each test item were used in determining the indicator of overconfidence in this study, its reliability could be compromised by factors influencing their ability to translate subjective judgements into numerical representations and motivation to do so honestly. Initiatives to assist them in the former were discussed earlier. In terms of motivational influences, Robson (2002) argues that subjects may feel that certain responses show them in a more favourable light, citing the example in educational research of a student aiming to please his/her tutor. This relates to the possible explanation for overconfidence discussed earlier in which, if they perceive social utility from doing so, respondents may feel motivated to express greater confidence than is warranted (Wright and Wishuda, 1982). Consequently, in this study, some may have been motivated to provide unjustifiably high confidence levels for self-presentation purposes (Camerer and Hogarth, 1999) in order to give the impression that they understood the issues on which they were being questioned. Alternatively, since the term overconfidence may be seen as implying arrogance or vanity (Yates, Lee and Shinotsuka, 1996), others may have tried to bias their responses in the direction of underconfidence, by providing ratings indicating inappropriately low levels of confidence. To help prevent these problems and reduce the potential for students providing responses intended to produce a desired outcome, while participants were made aware in advance that they would be participating in an activity designed to help them determine the extent of their learning on their study programme,

they were not informed that the specific focus was on self-monitoring or overconfidence in knowledge. During data collection, the researcher also verbally assured them that the results of the activity would have no direct bearing on summative assessment of their performance on their study programme. While student registration codes were collected to permit subsequent investigation of the association between metaknowledge and academic performance, it was also emphasised, both in writing on the questionnaire instructions (see Appendix 8) and verbally, that results would be anonymised and aggregated for research purposes.

3.4.5.2 Observer Bias

Observer bias occurs when the researcher does not make objective judgements in the data collection process. In this study, this potential problem was mitigated in that primary data was collected using a series of test questions, to which there was only one correct response, as verified by an academic member of staff with expertise in the knowledge domain in question. Expressions of confidence were also elicited using an objective indicator, using a continuous scale ranging between 25% and 100%. Consequently, the primary data collection procedure facilitated objective interpretation of participant responses. The secondary data, in respect of participants' academic performance on their study programme, was based on subjective judgements made by university tutors during the assessment process and was therefore subject to potential observer bias. However, to mitigate this, as well as using the assessment criteria discussed earlier to guide assessors, the university also has institutional procedures in place to moderate their judgements in the assessment process and thus reduce the potential for observer bias and promote consistency. These include moderation of marks by other internal members of staff and in the case of all work above first year undergraduate level, external moderation by suitably qualified assessors from outside the university.

3.4.5.3 Situational Factors

When using a test instrument in research, as is the case for this study, reliability may be affected by the conditions in which it is employed (Cohen, Manion and Morrison,

2000). Consequently, data collection events were all conducted in a consistent manner as described earlier. Each took place in the teaching room which participants regularly used for the study module on which the test was based. As this accommodation was of a similar design for all groups, physical conditions and contexts were consistent. Another situational factor in the case of a test is its perceived importance (Cohen, Manion and Morrison, 2000) and consequently, participants were given consistent information, which indicated that completion of the research instrument would have no impact on their summatively assessed performance on their study programme and that results would be anonymised and aggregated. Another potential issue is that in some cases respondents were completing a test which was not in their first language. However, while this had the potential to affect the reliability of the study, this was tempered by the requirement for such students to demonstrate their competence in English as part of the university application and admissions process. While this would not have been sufficient to ensure equal English language proficiency among participants, the impact of this would have been confined to their understanding of the task requirements, which were clarified as explained earlier to attempt to mitigate this, rather than the test questions themselves. This is because in cases where they were uncertain about their understanding of a question, then provided they understood the nature of the task, this uncertainty could be taken into account when making the associated confidence judgement.

3.4.6 Validity

Validity relates to the 'integrity of the conclusions that are generated from a piece of research' (Bryman, 2008 p.32) and different types of validity are typically distinguished. These relate to measurement, ecological, internal and external validity and each of these is discussed below in the context of this study.

3.4.6.1 Measurement Validity

Measurement validity, which is often called construct validity, is concerned with whether the measure or indicator actually reflects the concept for which it has been devised (Bryman, 2008). It is important to consider this as well as the reliability of a

measure, since it is of limited use if it provides results which are stable, but do not relate to the concept in question. In this study, overconfidence in knowledge was measured using the bias score and academic competence, using participants' assessment performance scores on their study programme. Therefore, the question is whether these adequately reflect the concepts in question. Robson (2002) argues that, because of the complexities surrounding them, most ways of collecting data are likely to be limited to some extent and contends that in many investigations, the reasonableness of the validity of a particular measure can, to some extent, be assessed intuitively, a process which establishes its face validity. The main determinant of validity is the extent to which the measure corresponds with the theoretical definition of the concept in question (Muijs, 2004). In this case, overconfidence in knowledge has been defined as the unjustifiable belief that one's judgements are accurate (Lichtenstein, Fischhoff and Phillips, 1982). Since the bias score is determined by comparing respondents' expressions of their belief in the accuracy of their judgements (i.e. their confidence estimates) with their actual accuracy (i.e. the knowledge score), it reflects this definition and enables the classification of individuals as overconfident, underconfident or neither. It also provides a means of indicating the extent of judgemental bias since, for example, it classifies those providing high confidence levels, but achieving low knowledge scores, as highly overconfident.

The other concept addressed in the study is academic performance and this is determined using indicators based on tutor assessments during participants' study programme. While these are subjective to some extent, their validity is supported by the fact that they were produced by an established and credible institution of higher education, which has procedures to promote valid assessment of academic performance.

3.4.6.2 Ecological Validity

A discussed earlier, ecological validity relates to the extent to which individuals' natural behaviour is reflected in experimental scenarios (Bem and Lord, 1979) and whether findings from social research are applicable in natural settings (Bryman, 2008). It has been argued that subjects in real life settings do not display the same cognitive limitations as those in laboratory settings (Christensen-Szalanski and Bushyhead, 1981)

and suggested that overconfidence detected in some of the previous studies reported may not be ecologically valid. Hacker, Bol and Bahbahani (2008) are critical of the fact that very few calibration studies have been undertaken in a classroom environment. Additionally, since many previous studies have focused on general knowledge, Stone (2000) suggests that further research on more specific subject matter is needed and Lundeberg, Fox, and Punccohar (1994) have encouraged more investigation of individual differences in overconfidence in natural settings, such as when answering tests or sitting examinations. This study addresses these issues by conducting the research in a classroom environment, based on knowledge related to the study programme of participants. Ehrlinger et al (2008 p.101) argued that using 'naturally occurring agents', such as a course instructor rather than an experimenter, in devising tasks for student participants in confidence judgements, would enhance the ecological validity of the design. In this study, this was taken into account in that tests were compiled by tutors with expertise in the subject matter in question.

3.4.6.3 Internal Validity

Internal validity is concerned with whether conclusions regarding causal links between variables can be supported. When addressing causality, a distinction is made between independent variables, which have the causal impact, and dependent variables, which they influence (Bryman, 2008) and internal validity relates the extent to which changes in dependent variables can be attributed to independent variables (Sarantakos, 2005). Robson (2002) explains that an experimental approach is required to investigate this, but argues that since there are many factors which threaten internal validity, a belief that these can be eliminated in social research by adopting appropriate methods is consistent with a discredited positivist approach. Acknowledging the potential implications of such threats on the other hand complies with a post-positivist approach, in which all methods are deemed fallible. True experiments may also be impractical in social research as manipulation of variables or assigning subjects to control groups may be either unfeasible or unethical (Gill and Johnson, 1991). Consequently, while correlational relationships in quantitative studies such as this are important, they do not necessarily demonstrate causality (Sternberg, 1998), nor in their own right, help to understand what lies behind the relationship Robson (2002). Nash (1999 p.109) emphasises this by

distinguishing between ‘causal correlations and contingent correlations’, explaining that in some cases, it is necessary to highlight that associations between data may not necessarily represent causal relationships. This study is a case in point, as the association between confidence bias and academic performance is explored using a non-experimental method. It does not entail manipulating variables or the use of control groups and while this approach does not permit claims of causality, Hacker, Bol and Bahbahani (2008) argue that it may be the best method of investigating calibration in a natural setting. Kemp and Holmwood (2003) highlight the value of such studies, arguing that while statistical analysis may not in itself establish causality, looking for patterns and associations can play an important role in the identification of causes of events in social science. Consequently, it is important to emphasise that this study does not seek, or claim to establish, causal links between overconfidence bias and academic performance but rather, aims to determine whether associations exists, which may be indicative of potential causes.

3.4.6.4 External Validity

External validity is concerned with the extent to which findings from research are potentially applicable to the wider population in question (Cohen, Manion and Morrison, 2000). Sarantakos (2005) explains that a problem which may threaten external validity arises where some subjects, as a result of being chosen to participate in the study, are motivated to find out more about the issue under investigation and therefore become more knowledgeable about it than an average member of the population. To help overcome this problem in this study, subjects were not made aware of the explicit focus of the research before completing the test. Additionally, any student who participated in more than one data collection event, due to the specific nature of their study programme, was included only on the basis of their first submission, to ensure that data was collected from participants who were unaware of the specific issue under investigation in the study.

Sampling

Another issue of concern in external validity is sampling, which is the choice of a number of subjects, from whom data is collected in order to draw conclusions about the population represented (Jankowicz, 1995). In some cases, it is possible to conduct a census in which data is collected relating to the entire population under investigation. While this may provide a more complete picture than sampling, it can be very costly and time consuming and using samples can therefore provide reliable results more efficiently (Davis, 2005). However, inadequate or biased sampling of the population can result in an unrepresentative sample, though it is very difficult to obtain an entirely representative sample (Bryman, 2008) and interpretation of results should therefore take into account the sampling procedures employed. Cohen et al (2000) suggest that there are four key factors in sampling:

1) Sample size
2) Representativeness and parameters of the sample
3) The sampling strategy to be employed
4) Access to the sample

Sample Size

Pring (2000) suggests that conclusions drawn from quantitative research can only be tentative, but that the larger the sample, the more confidence one can have in those conclusions. However, due to the limited resources typically available for research, determining the sample size requires a balance between the cost of obtaining data and statistical efficiency (Davis, 2005). Cohen, Manion and Morrison (2000) suggest that for random probability sampling, sample size can be determined by either exercising prudence and ensuring that the sample represents the wider feature of the population by including a minimum number of cases in each sub-group, or by referring to mathematical tables to determine the minimum number required in the sample to represent the population. Table 4 provides an example which indicates the sample size required for different population sizes at confidence levels of 95% and 99%. It is the

absolute sample size which is most important rather than the size relative to the population (Davis, 2005), as can be seen in the table, where the required sample size relative to the population falls as the population size increases.

Table 4 - Sample Size, Sampling Errors and Confidence Levels

	Sampling Error / (Confidence Level)			**Sampling Error / (Confidence Level)**	
	5% / (95%)			**1% / (99%)**	
	Sample			***Sample***	
Total Population Size (N)	Size (S)	Sample as % of Population		Sample Size (S)	Sample as % of Population
50	44	88.0%		50	100.0%
100	79	79.0%		99	99.0%
200	132	66.0%		196	98.0%
500	217	43.4%		476	95.2%
1,000	278	28.8%		907	90.7%
2,000	322	16.1%		1,661	83.1%
5,000	357	7.1%		3.311	66.2%
10,000	370	3.7%		4,950	49.5%
20,000	377	1.9%		6,578	32.9%
50,000	381	0.8%		8,195	16.4%
100,000	383	0.4%		8,926	8.9%
1,000,000	384	0.1%		9,706	1.0%

Source: Adapted from Cohen, Manion and Morrison (2000 p.95)

In this study, the population was all full time students studying business at the university in question and since the data was collected in a period which straddled two academic years, the mean of these two periods was used to determine this. This indicated a total of 2,492 students and, in accordance with the table this requires a sample of between 322 and 357 students at a 95% confidence level. Consequently, a sample of at least 357 participants was sought.

Representativeness and Parameters of the Sample

Bryman (2008) explains that it is usually accepted that representativeness will result from the use of a probability sample, where selection is made in a random manner, in which each unit of the population has a known chance of being chosen. However, he argues that in many, if not most, cases it may not be possible to use random sampling and that, in those circumstances, it is important to explain clearly the approach that has been taken and the reasons for that. Limited resources may necessitate the use of a sampling approach that is less than ideal (Baker, 1999) and since this was the case for this study, the manner in which the sample was chosen is discussed below.

Sampling Strategy

To address the concerns about ecological validity discussed earlier, the knowledge domain addressed in this research study related to students' study programmes and as the population was enrolled on various programmes, the nature of this knowledge differed accordingly. Since random sampling would have resulted in participants being drawn from a wide variety of study programmes, this would have required the design of a large number of different tests, to ensure that every participant was tested on knowledge related to their study programme. Consequently, as this would have been prohibitively time consuming and expensive, random sampling was rejected and a cluster sampling approach employed.

Cluster sampling is mainly used where a sampling frame of all participants is either unavailable or unsuitable (Sarantakos, 2005). It entails dividing the population into a number of clusters and drawing respondents from a limited number of these. This approach is commonly used in small-scale research (Cohen, Manion and Morrison, 2000) and the main reason for its popularity is its efficiency (Davis, 2005). It is therefore particularly useful in circumstances where it is difficult, or expensive, to obtain data by sampling from the whole population (Robson, 2002), as is the case in this study. The sample can be drawn from the chosen clusters either randomly or, as is often the case in educational research, by targeting all within the cluster (Muijs, 2004). The clusters used in the study were eight study modules, which operate across four levels of

study, levels 4-6 relating to undergraduates and level 7 to postgraduates. As discussed earlier, each related to financial and quantitative aspects of participants' study programmes. To ensure a wider representation of participants, the modules chosen for each study level included one studied only by students specialising in accounting or finance and another by those who were not. Sampling in this manner permitted the investigation of overconfidence bias in the context of participants' subject-related knowledge, by tailoring the research instrument to reflect the relevant knowledge domain for each cluster. This approach also gave the opportunity to integrate the data collection activity into the study programme of the modules in question, which allowed all students attending these sessions to be targeted, rather than randomly selecting a sub-sample from each. As well as enhancing its ecological validity, conducting the research in a natural setting in this way was therefore also designed to assist in increasing the sample size, by accessing potential participants at a time when it was easier to hold their attention as advocated by Loftus and Wagenaar (1988) in the study of overconfidence in lawyers discussed earlier. While this approach to increasing the sample size is likely to enhance the representativeness of the study in one respect, it must be acknowledged that it may be limited by the fact that sampling was not random. However Schwab (1985), in highlighting the sampling problems in social research, explains that random sampling in organisational studies is rare. He argues that almost all empirical studies in this field use convenience samples and that insistence on random sampling, to facilitate generalisation of results, would result in the rejection of most research submitted to academic journals for publication. Nevertheless, when using cluster sampling it is necessary to acknowledge the implications for generalisability (Cohen, Manion and Morrison, 2000) and in comparison with random sampling, this method will tend to produce larger errors for comparable sample sizes (Davis, 2005).

The extent to which the sample is to be divided into sub-groups also influences the size and nature of the sample (Ticehurst and Veal, 1999). However, the more factors included, the more complicated the sampling process becomes and therefore Cohen, Manion and Morrison (2000) suggest aiming for simplicity when deciding on sub-groups for analysis. Robson (2002) cites Mertens' (1998) suggested rule of thumb, for non-experimental relational designs, of approximately 15 participants per variable and Borg and Gall's (1979) suggestion for surveys of approximately 100 observations for

major sub-groups investigated and 20-50 for minor sub-groups. Consequently, in this study, since the main sub-groups of interest related to age, gender and country of origin (UK and Chinese), a target of approximately 100 cases for each of these was sought by sampling a sufficiently high number of participants. However, it was acknowledged that in the case of age, due to the relative homogeneity of students studying at the university in this respect, it may be necessary to re-classify respondents into a smaller number of sub-groupings to achieve this.

While the procedure described above was used to select a sample of students, the number providing usable data was reduced by two factors. Firstly, in accordance with ethical research procedures, each was asked to provide written consent to using their responses for research purposes and those who did not were excluded. Secondly, respondents were also excluded from the sample on the grounds of inappropriate use of the confidence scale, where they provided confidence levels below 25% or responded inappropriately to the diagnostic 'British Airways' test question designed for this purpose.

Access to the Sample

It is also important to ensure that access to the research subjects is both permitted and practicable (Cohen, Manion and Morrison, 2000). Regarding permission, ethical approval was obtained from the Research Ethics Committee at the university. To gain access, the researcher also discussed the nature of the study with tutors responsible for the learning of the students in question, explaining the potential benefits of their participation, through feedback they would be provided with after completing the task. This resulted in agreement to integrate the data collection activity into the study programmes in question as explained earlier and thus ensured access to the sample.

3.5 Data Entry and Analysis

When designing research, it is important to consider how data collected may be analysed and the techniques employed will depend on the research method adopted (Denscombe, 2003). In this study, two types of software were used for data analysis.

Initially, data from completed questionnaires was transferred to Microsoft Excel spreadsheets, which were used to determine mean confidence, the mean knowledge score, and the bias score for each respondent. They were also designed to identify which respondents had provided usable data, by indicating whether they had agreed to participate in the study and had used the confidence scale appropriately. The latter was determined by ascertaining whether confidence ratings below 25% had been used and whether a 25% rating had been provided for the diagnostic 'British Airways' question used to detect inappropriate use of the scale. Secondary data, in respect of the academic performance of participants on their study programme, which was obtained from the university's marks recording system, was also recorded on the spreadsheet. This spreadsheet data was subsequently imported into an SPSS statistical software package for subsequent data analysis and to facilitate this analysis, a coding plan was devised. Codes allow subsequent processing on a computer (Bryman, 2008) and permit the identification of patterns among the variables addressed in the questionnaire (Czaja and Blair, 1996) and those used in this study are shown in Appendix 9.

Statistical tests were undertaken to investigate associations and differences between sub-groups as appropriate and these tested for statistical significance at the 5% level. This is commonly used in social research and indicates that there is a less than 5% chance of the relationship found in the sample arising by chance only and not occurring in the population (Bryman, 2008). Where appropriate, results indicating a 1% significance level were also highlighted. The specific statistical tests undertaken are discussed below.

3.5.1 Task Difficulty Effect

Since previous studies have reported that task difficulty may influence overconfidence, it was necessary to ensure that the eight knowledge based tests used in the study did not significantly differ in terms of their difficulty. To facilitate this, each test was approved by the relevant study module tutor as being a reasonable test of the respondents' knowledge, in the context of their level of study and the knowledge domain being tested. However, having attempted to design tests of equal difficulty, this should subsequently be verified statistically (Yates, Lee and Shinotsuka, 1996) and the

knowledge score was used for this purpose, since it represents the proportion of questions answered correctly by each respondent. Hacker et al (2000) compared the median knowledge score across each of three confidence tests they employed and concluded that since these were equal, the potential confounding effects of using different tests were minimised. Grimes (2002) and Hacker, Bol and Bahbahani (2008) meanwhile tested for differences in mean knowledge scores for each test and this approach was also adopted in this study. When investigating differences in means between more than two groups, as in this case where there are eight, ANOVA, which tests differences in three or more means (Tabachnick and Fidel 2001), can be used. Different types of ANOVA test can be employed, depending on the whether the independent variables are independent measures, in which respondents fall into only one condition, or repeated measures, in which each respondent has a score in each condition of the variable (Hinton et al 2004). In this case, each of the independent variables is an independent measure (i.e. each participant had a knowledge score for only one test) and therefore, independent measures ANOVA was appropriate.

An indication of significantly different levels of difficulty across the eight tests would require an intervention to control for this. One approach for doing so, which was used by Klayman et al (1999), would be to use relative overconfidence. This indicator removes differences in difficulty caused by the use of different tests, by determining the bias score for each participant relative to that of others taking the same test, as follows:

$$\textit{Relative overconfidence} = \textit{Bias score} - \textit{Mean bias score for all participants taking that test}$$

Since the bias score for each participant is compared with the mean for others taking the same test, employing this method would result in mean relative overconfidence scores of zero for each of the eight groups of participants completing the different tests used. This would facilitate investigation of individual differences where participants in each condition were drawn from across all eight tests, such as gender for example, where each test included both male and female participants. However, it would hamper any investigation of students by level of study, since at each study level only two different tests were used.

Consequently, an alternative and less restrictive approach, which reflects that employed by Yates et al (1989), was adopted. This entails deriving, for each of the eight tests used in the study, a sub-set of questions which did not significantly differ in difficulty across the eight tests. This could be achieved for each test by initially determining the mean correct response rate for each of the 30 original questions used. These can be subsequently randomly sampled, using SPSS software, to produce subsets of 25 questions for each test, each with a mean knowledge score approximately equal to that for the entire sample of participants across all eight tests. This procedure therefore permits the design of eight tests of approximately equivalent difficulty in order to control for the task difficulty effect.

3.5.2 Investigating Overconfidence

The hypotheses established in Chapter 2 are shown in Table 5. The first of these was tested using a one sample t-test on bias scores, using a test value of zero, since a positive bias score indicates overconfidence.

Table 5 – Hypotheses Tested in the Study

Hypothesis
H_1: Students are overconfident in their knowledge
H_2: Overconfidence differs between older and younger students
H_3: Overconfidence is greater for male students
H_4: Overconfidence is greater for Chinese than UK students
H_5: There is a negative association between overconfidence and academic performance

3.5.3 Individual Differences

The study also aims to explore individual differences in overconfidence in respect of gender, age, nationality and level of study. To investigate differences in the mean bias score between two groups, as in the case of gender for example, independent samples t-tests were used. In the case of a directional hypothesis (e.g. overconfidence is greater for male students) a one tailed test was used. When analysing differences between groups, the problem of differential selection, in which members of the group differ in respects other than the explanatory variable in question, may occur. Robson (2002) suggests that, if random allocation to groups is not possible, this can be addressed through analysis of subgroups. This was achieved by investigating differences in two dimensions, using two-factor ANOVA, to investigate the potential moderating effect of a second variable, such as whether there are gender differences in overconfidence bias between students from China and the UK.

The use of t-tests and ANOVA to explore differences in means assumes that the analysis is conducted on interval data, which is representative of the population and is normally distributed. However, should the data not be normally distributed, results can still be meaningfully interpreted, provided that samples of over 30 are used (Hinton et al, 2004). In this case, interval data was used, as well as samples in excess of 30 for respondents as a whole and for each of the sub-groups analysed in respect of age, gender and country of origin. In terms of the extent to which the sample was representative of the population, Muijs (2004) explains that, as is the case in this study, it is often difficult to conduct random sampling in educational studies and that research has indicated that t-tests are robust to violation of this assumption, provided large samples are used. As well as using a large sample for the study, representativeness was also addressed by investigating the extent to which it represented the population in respect of the individual characteristics investigated; age, gender and country of origin. T-tests and ANOVA also assume that samples are drawn from populations with equal variances (the homogeneity of variance assumption). Violation of this when performing t-tests to compare two groups required using the result generated by the test in which equal variances are not assumed (Hinton, 2004) and when investigating differences between more than two groups, necessitated the use of the non-parametric Kruskal-Wallis test, as the analogous test to the ANOVA explained above.

3.5.4 Overconfidence and Academic Performance

The association between overconfidence and academic performance was analysed using Pearson's correlation between the participants' bias scores and overall marks for the level of study in which the confidence data was collected, since each indicator comprised interval data (Hinton et al, 2004). Since the hypothesis being tested was directional, a one tailed test was appropriate. This approach was also used to investigate the association between overconfidence and academic performance in each of the various different mode of assessment used in the participants' study programmes (examinations, coursework etc.).

3.6 Ethical Issues

Research should be conducted in an ethical manner and in order to protect those involved in the study, social researchers should behave honestly and with integrity and respect the rights and dignity of the participants, ensuring that they are not harmed as a result of their participation (Denscombe, 2003). These principles are incorporated in the university ethics policy (University of Northumbria, 2009a), which embraces the principles of beneficence, implying that the research should have a conceivable benefit and nonmaleficence, which indicates that it should not be harmful (Rosenthal and Rosnow, 2008). The policy therefore requires that the anticipated consequences of the work, and well being of others, should be taken into account when conducting research. Researchers should also obtain the consent of those who may be willing to participate in studies (Reeves and Harper, 1981) and this is particularly important in cases where they may be in a position of relative power, as is the case in this study, where the researcher is an academic member of staff in the institution in which the research was conducted. This requirement is also made clear in the university ethics policy, which explains that the participation of human participants in research should be on the basis of informed consent. This can be implicit, through respondents returning a questionnaire (Fisher, 2004) and the university policy for informed consent in research (University of Northumbria, 2009b p.3) accepts this unless 'sensitive personal data as defined by the Data Protection Act' is collected. This definition includes information regarding 'the

racial or ethnic origin of the data subject' (Great Britain, Data Protection Act, 1998 p.2). While these were not specifically requested in this study, respondents were asked to record their country of origin, and therefore they were asked to indicate in writing on the questionnaire whether they consented to their data being used in the study. It was emphasised that they were under no obligation, and should not feel pressured, to do so. Data in respect of those who did not consent to participation was not used in the research. However, along with those who did consent, they were provided with feedback on the results of the activity. This included their knowledge and bias scores and therefore, they were provided with both performance and cognitive feedback, which provided a more complete assessment of their knowledge, in that they were made aware of not only what they knew, but also what they thought they knew (Renner and Renner, 2001). They were also provided with information to assist in the interpretation of their results, to allow them to reflect on their potential implications for their future learning. As well as addressing ethical issues, this was designed to assist in increasing the sample size and enhance the reliability of the data collected, as providing information to respondents in this way can motivate them to participate and provide considered responses (Davis, 2005). Boud (1995) advises that public presentations of self-assessments should be avoided, unless initiatives are in place to protect respondents' self-esteem. Consequently, each student was provided with a unique reference number, which was shown on their questionnaire and known only to them. They were then able to use this number to access their results, from information made available via the e-learning platform used in the university.

The university policy for informed consent in research (University of Northumbria, 2009b p.4) also requires that 'all data must be encoded or anonymised if possible' and this was achieved through the use of individual student codes and the aggregation of data for analysis. Consequently, in compliance with the principles of nonmaleficence, those taking part were protected from harm as a result of their participation in the study. Additionally, in terms of beneficence, there were potential benefits from the research both for participants in the study, through personal feedback they received after completing the research instrument, as well as for future learners who may benefit from the manner in which the findings emerging from the study may inform the development of future learning activities.

While an open approach is usually considered more ethical (Maykut and Morehouse, 1994), disclosing the specific research objective may bias the responses provided (Gill and Johnson, 1991). In these circumstances, it may be appropriate to disclose the nature, though not the precise aim, of the work, if those being studied are not harmed as a result of their participation. Consequently, while they were informed that the intention was to use the data for research purposes, to assist in understanding influences on student learning, participants were not made aware that the study specifically addressed overconfidence in knowledge, in order to prevent this information biasing their responses. However, as discussed above, after completing the questionnaire they were provided with feedback, which included their bias score and highlighted the consequences of poor knowledge monitoring.

3.7 Conclusion

This chapter has considered the philosophical foundations underpinning the research and the design used to answer the research questions developed in Chapter 2. It is informed by realist ontology, empiricist epistemology and quantitative methodology and entails a deductive process, in which overconfidence is operationalised and theory tested empirically, in a large business school based in a UK university. However, while adopting the stance that an objective reality exists which may be investigated empirically, it is acknowledged that it may not be possible to discover it with certainty, due to limitations in the manner in which the research was conducted. The difficulties in operationalising social concepts such as overconfidence in knowledge are recognised, as well as the potential impact of bias on the part of both the participants and the researcher. Consequently, in order that readers may take limitations of the study into account when interpreting its findings, the research design has been fully explained.

The study employs a non-experimental, fixed design using a research instrument comprising a closed response, two part questionnaire, to collect demographic details of participants as well as gathering data which can be used to generate an indicator reflecting knowledge monitoring accuracy. In terms of its ecological validity, the research design responds to calls for more studies using tests in a classroom environment (Lundeberg, Fox, and Punccohar, 1994; Hacker, Bol and Bahbahani, 2008)

and suggestions that research should focus on rich, specific subject matter gained over a period of time (Stone, 2000; Hacker et al, 2000), by investigating knowledge monitoring in the context of the study programmes of participants. Integrating data collection events into the participants' study programmes, was intended not only to conduct the study in a natural setting, but also assist in increasing the sample size. Basing the study on participants' knowledge related to their study programme requires the use of cluster sampling, due to the difficulties and resource implications of random sampling in a large institution with students following a wide variety of study programmes. Consequently, the approach adopted represents a trade off between two types of validity in that the strategy employed to enhance ecological validity means that, in common with many other organisational studies, random sampling was not used. While this has potential consequences for the representativeness of the sample, the clusters used ensured that participants were drawn from all three levels of undergraduate study, as well as postgraduate programmes.

The research instrument incorporates a test comprising 30 multiple-choice questions, which was used to collect data which can be used to generate indicators of knowledge monitoring accuracy. Each question required respondents to indicate their choice of the correct answer, as well as a confidence judgement, using a graduated scale in the range 25%-100%. In terms of its relationship with the conceptual framework and dimensions of calibration discussed in Chapter 2, the study is therefore an investigation of metacognitive monitoring using retrospective, local meta-level judgements, in which respondents provide postdictions of accuracy in respect of individual test items. The bias score used in many previous studies was employed as an indicator of overconfidence and in order to enhance its reliability, a mechanism to identify inappropriate use of the confidence scale was incorporated in the research instrument. While it is acknowledged that this may not necessarily eliminate all difficulties respondents may have in translating subjective judgements into quantitative measures, it should enhance the reliability of data used in the study, by excluding those who provide evidence suggesting they have done so inappropriately. Similarly, while it is acknowledged that responses from participants could be biased by the social utility some may attach to confidence judgements, the research design includes initiatives to mitigate these. These include discussions with students regarding the manner in which

data collected is processed and feedback provided. To permit investigation of the association between overconfidence and academic performance, the latter was operationalised using assessment results from the participants' study programme. This approach addresses problems highlighted in the literature associated with using the same research instrument to operationalise both overconfidence (the bias score) and competence (the knowledge score). This secondary data, related to the participants' assessment results on their study programme, was gathered from the institution's marks recording system and as well as their overall performance, included sub-component marks related to various modes of assessment used at the institution. While these indicators arise from subjective judgements made by academic staff, the institution has procedures in place to moderate these assessments to promote consistency in the manner in which they are made.

Since the study entailed collecting data from individuals, the ethical issues associated with doing so were addressed. In order to ensure that respondents were not harmed through their participation, their anonymity was protected and data aggregated for analysis. The questionnaire asked them to indicate in writing whether they consented to their responses being used in the research and where they declined, their data was excluded. While being informed that the purpose of the research was to investigate issues associated with student learning, they were not informed of the specific emphasis on overconfidence in knowledge, in order to prevent this information influencing their responses. However, shortly after completing the test, they were provided with individual feedback of their results, as well as information to assist them in interpreting this.

Having addressed the philosophical foundations of the research and methodology adopted and discussed the research design in detail, in this chapter, the results arising from deploying this design to answer the research questions developed in Chapter 2 are considered in the next chapter.

4. Findings and Discussion

4.1 Introduction

This chapter reports the findings arising from the application of the research design explained in Chapter 3 to the research questions developed in Chapter 2. Their relationship with those reported in previous studies discussed in Chapter 2 is also addressed as well as their potential implications. Initially, the participants in the sample are analysed, indicating the extent to which they are representative of the population from which they were drawn. The procedure designed to control for task difficulty is then discussed to clarify how this was implemented before reporting findings in respect of knowledge monitoring accuracy. This initially addresses the extent to which overconfidence in knowledge was detected in participants generally and its association with both knowledge and confidence. Individual differences are subsequently analysed, before considering the association between overconfidence and various aspects of academic performance.

4.2 Participants

While questionnaires were completed by 606 participants, as shown in Table 6, these were not all used in the study.

Table 6 - Determination of the Final Sample

	Students
Students completing the test	606
Less: those completing twice	(15)
	591
Less: not consenting to use of data	(26)
Consenting respondents	565
Less: using confidence scale inappropriately	(57)
Final sample	508

Despite efforts to prevent this, by choosing non-overlapping cluster samples, 15 of those completing the test attended two data collection sessions and therefore submitted two questionnaires, each based on a different knowledge domain. However, only the first of these was included in each case, since having completed it once and received feedback, this may have influenced responses provided in their second attempt. Of the remaining 591 participants, 26 (4%) indicated that they did not wish to their results to be used in the study and were therefore excluded. This high consent rate suggests that initiatives designed to enhance it, including reassuring students about how the data would be used and explaining that individual feedback would be provided, were effective. This left 565 respondents, of which three were excluded from the study on the grounds of using confidence levels below 25%, and a further 54 (10%) for not providing a confidence level of 25% for the diagnostic 'British Airways' question designed to test for inappropriate use of the confidence scale. As shown in Table 7, analysis of these indicated that while two did not disclose their country of origin, 13 were UK students (comprising 5% of total UK students) and 42 were non-UK (15% of total non-UK). Thus proportionately fewer UK students were excluded on the basis of their inappropriate use of the confidence scale. The remaining 508 participants comprised the final sample analysed for the study.

Table 7 – Analysis of Respondents Excluded for Inappropriate Use of the Confidence Scale

Country of Origin	*Consenting Respondents*	*Inappropriately Using Confidence Scale*	
	n	*n*	*% of Consenting Respondents*
UK	282	13	5%
Non-UK	278	42	15%
	560	55	
Not disclosed	5	2	
Total	565	57	10%

4.2.1 Individual Characteristics

4.2.1.1 Gender

Three participants did not disclose their gender and the remainder are detailed in Table 8, along with the gender split in the population of full time business students studying at the university. This was determined using the mean proportions for 2006/07 and 2007/08, the period during which the data was collected for the study. The table shows the proportion of males in the sample exceeding that in the population by 2%, with females showing a corresponding deficit. It therefore indicates that the sample closely represented the population in respect of gender.

Table 8 – Analysis of Participants by Gender

	Participant Sample (S)		*Population (P)*		*(S-P)*
	N	*%*	*N*	*%*	*%*
Male	281	56%	1,339	54%	+2%
Female	224	44%	1,153	46%	-2%
Total	505	100%	2,492	100%	
Did not disclose	3				
Total sample	508				

4.2.1.2 Age

All except four respondents disclosed their age group and, as shown in Table 9, the majority (92%) were aged between 18 and 25 years. Consequently, for the purposes of analysis, the older age bands were collapsed as shown in Table 10. This indicates that the proportions in each age group in the sample were within 1% of that in the population and consequently, that it was representative in this respect. However, the relatively narrow age distribution does constrain the investigation of age differences to some extent.

Table 9 – Analysis of Participants by Age Group

	Participant Sample (S)		
	N	*%*	*Cum. %*
18-21 years	297	59%	59%
22-25 years	166	33%	92%
26-29 years	26	5%	97%
30-33 years	8	2%	99%
34+ years	7	1%	100%
Total	504	100%	
Did not disclose	4		
Total sample	508		

Table 10 – Regrouping of Participants by Age

	Participant Sample (S)		*Population (P)*		*(S-P)*
	N	*%*	*N*	*%*	*%*
18-21 years	297	59%	1,490	60%	-1%
22+ years	207	41%	1,002	40%	+1%
Total	504	100%	2,492	100%	
Did not disclose	4				
Total sample	508				

4.2.1.3 Country of Origin

Participants in the sample originated from 43 different countries and to permit comparison of the overconfidence levels of Chinese and UK students, the sample was sub-divided into these two groups and other students as shown in Table 11. This shows that while Chinese students are a little over, and UK under-represented, the sample does broadly reflect the population in this respect.

Table 11 – Analysis of Participants by Country of Origin

	Participant Sample (S)		*Population (P)*		*(S-P)*
	N	*%*	*N*	*%*	*%*
UK	269	53%	1,403	56%	-3%
China	102	20%	410	16%	4%
Other	134	27%	680	27%	0%
Total	505	100%	2,492	100%	
Did not disclose	3				
Total sample	508				

The participants analysed for the study therefore comprise a large sample of over 500 business students, which is broadly representative of the population of those studying at the institution in respect of gender, age group and country of origin. The sample also included at least 100 participants in each condition investigated in these major sub-groups, as recommended by Borg and Gall (1989).

4.3 Controlling for Task Difficulty

Given the potential confounding effect of task difficult in the investigation of overconfidence, this was tested for in the eight tests used in the study by analysing the differences in mean knowledge scores attained by participants in each, as these indicate the proportion of questions correctly answered. Since the data for knowledge scores violated the homogeneity of variances assumption implicit in an ANOVA test, a Kruskal-Wallis test (see Appendix 10) was used for this purpose and this indicated that there were significant differences in difficulty between the tests ($p=0.001<0.01$). Consequently, to control for this, eight tests of approximately equal difficulty were derived by using a sub-sample in each test of 25 questions from the original 30 used. This entailed a random sampling process to generate the 25 item tests, each with a mean knowledge score around the overall mean across all eight original 30 item tests (49.25%) and which did not differ significantly across each of the eight tests. As shown in Appendix 11, this was confirmed by a Kruskal-Wallis test ($p=0.980>0.05$). Data from

these refined tests were used in the study and from a potential maximum total of 12,700 confidence judgements (i.e. 508 respondents x 25 items), 12,564 responses were provided by participants, representing a completion rate of 99%.

4.4 Investigating Overconfidence

Table 12 shows measures of central tendency for mean confidence, knowledge score and bias score for all participants. The positive mean bias score of 8.9% indicates that, as a group, they displayed overconfidence and a Kolmogorov-Smirnov test (See Appendix 12) confirmed that it was normally distributed ($p=0.606>0.05$) and therefore satisfied this assumption associated with the use of parametric tests (Hinton et al, 2004).

Table 12 – Measures of Central Tendency for Confidence, Knowledge and Bias Score

	Mean Confidence %	*Knowledge Score %*	*Bias Score %*
Mean	59.1	50.2	8.9
Median	58.5	48.0	9.0
Mode	56.0	48.0	9.0
Std. Deviation	14.1	13.5	14.8

The following hypothesis was tested in respect of overconfidence:

H_0 = Students are not overconfident in their knowledge

H_1 = Students are overconfident in their knowledge

As shown in Appendix 13, a one sample t-test indicated that the difference between the mean bias score of 8.9% and a perfectly calibrated zero score was statistically significant at the 1% significance level ($p=0.001<0.01$). Consequently, the null

hypothesis is rejected and it is concluded that students are overconfident in their knowledge. Participants were also investigated using the standard used by Pallier (2003), in which scores between -5% and +5% are deemed indicative of good self-assessment. The results shown in Figure 21 indicate that 23% displayed good self-monitoring accuracy, with the majority (63%) demonstrating overconfidence and relatively few (14%), underconfidence.

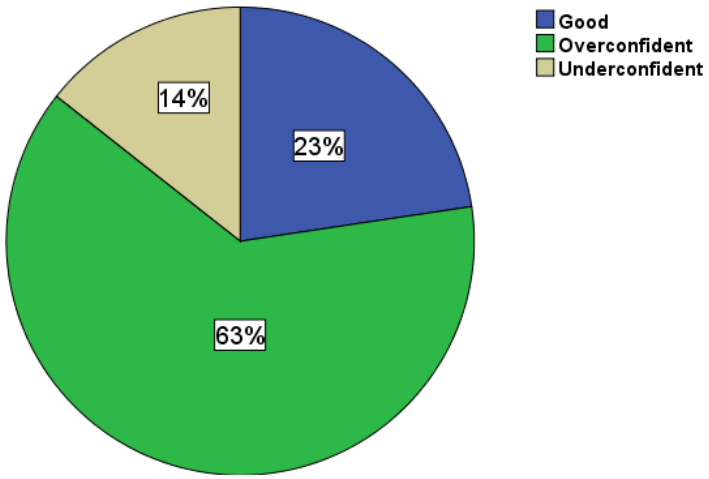


Figure 21: Self-Monitoring Accuracy of Participants

Figure 22 shows a calibration chart for all participants, plotting confidence against accuracy, and indicates that participants were particularly overconfident when expressing high levels of confidence. Further investigation of extreme levels of confidence indicated that respondents expressed 100% confidence in 23% of the total judgements made. However, as illustrated by the calibration curve, the accuracy rate for such responses fell well short of this at only 73%. Consequently, the inaccuracy rate of 27% (100%-73%) in respect of these 'false certainties', exceeds the 20% reported by Dunning et al (1990) and lies within the range of 17-30% in studies reported by Fischhoff, Slovic and Lichtenstein (1977).

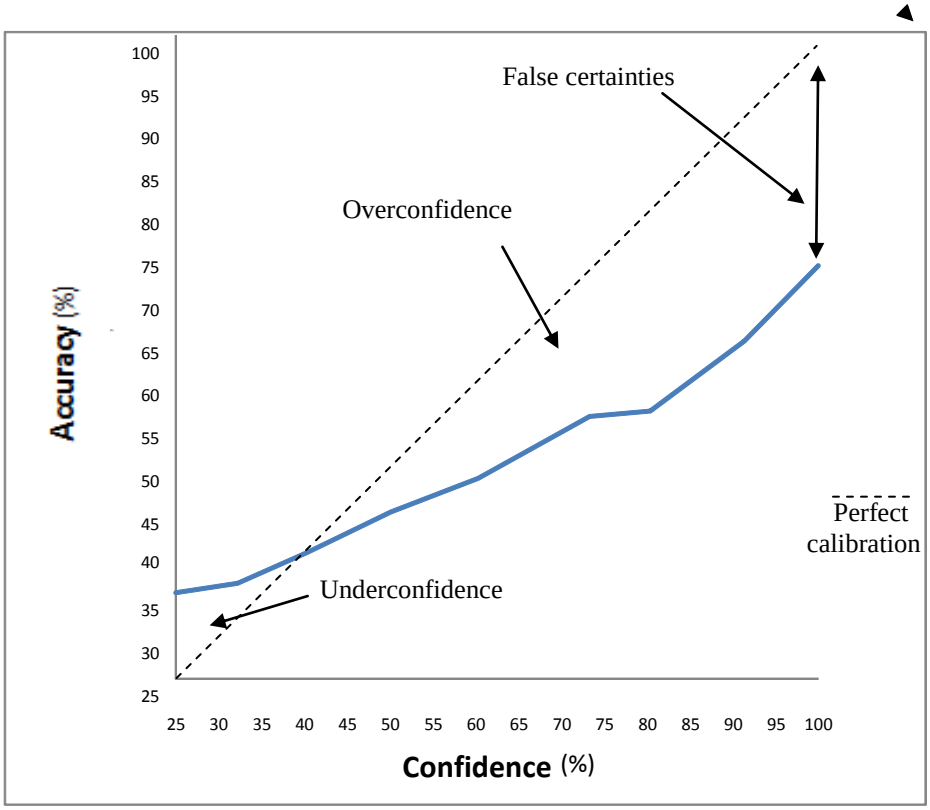


Figure 22: Calibration Curve – All Participants

4.4.1 Overconfidence and Knowledge

Investigation of the association between the bias score and knowledge score, as shown in Appendix 14, revealed moderately strong negative correlation, which was significant at the 1% level ($r=-0.504$, $p=0.001<0.01$). This indicates that those demonstrating greater overconfidence displayed less knowledge in the test. However, as discussed earlier, this result typically arises due to the method of investigation used. Using the same instrument to test for both knowledge and overconfidence tends to result in overconfidence for those displaying poor knowledge, due to the fact that it is more difficult to be underconfident when the proportion of questions answered correctly is low.

4.4.2 Overconfidence and Confidence

Previous studies have shown that confidence and overconfidence are positively related and in this case, the calibration chart in Figure 22 indicated greater overconfidence for

judgements in which participants expressed greater confidence. This was tested statistically and as shown in Appendix 15, indicated moderately strong positive correlation between confidence and overconfidence, which is significant at the 1% level ($r=+0.561$, $p=0.001<0.01$). Therefore it can be concluded that those who displayed more confidence also tended to demonstrate greater overconfidence, a finding which supports similar results reported by Klayman et al (1999) and Loftus and Wagenaar (1988). Again, it can be argued that the use of the same instrument to determine indicators for both confidence and overconfidence contributes to this finding, since it is more difficult for those providing relatively low confidence judgements to demonstrate overconfidence.

4.4.3 Individual Differences

Having determined that participants displayed a general tendency for overconfidence, individual differences were investigated to determine its association with age, gender and country of origin.

4.4.3.1 Age

The following hypothesis was tested with respect to age:

H_0 = No age differences in overconfidence exist

H_2 = Overconfidence differs between older and younger students

Since the age group classifications were compressed to two groups, this was achieved by investigating whether overconfidence differed between the younger group, aged 18-21 years, and those aged 22 years and above. Table 13 shows that the older group displayed more confidence in their responses, but this was not justified by their performance, since their mean knowledge score was very close to that for the younger group. Consequently, while both groups displayed overconfidence, it was evident to a greater extent in the older group, with a bias score of 11.3% as compared with 7.3% for

younger students. An independent samples t-test (see Appendix 16) indicated that this difference was statistically significant at the 1% level ($p=0.003<0.01$). The null hypothesis is therefore rejected and the results indicate greater overconfidence for older students.

Table 13 – Mean Confidence, Knowledge and Bias Score by Age Group

	Mean Confidence %	*Knowledge Score %*	*Bias Score %*
18-21 years	57.6	50.3	7.3
22+ years	61.3	50.0	11.3
Total	59.1	50.2	8.9

This finding does not support earlier studies which reported the tendency for overconfidence to reduce with age (Grimes, 2002; Fischhoff, 1992), nor Fitzgerald, White and Gruppen's (2003) suggestion that self-assessment ability is mainly learned in childhood and once in adulthood tends to be fixed. However, it is acknowledged that the extent to which the study was able to investigate the influence of age was limited to some extent, due to the narrow age range of the population.

The calibration curves shown in Figure 23 indicate a similar pattern for each age group in respect of accuracy at different levels of confidence. In terms of questions which produced extremely high levels of confidence, accuracy for those in which 100% confidence was expressed was similar for each group, with 18-21 year olds answering 74% of such questions correctly and the older group 72%. However, the younger group expressed this degree of confidence more frequently, doing so for 26% of questions, as compared with 21% for the older group. This outcome is interesting in view of Fischhoff and McGregor's (1982) finding of more accurate self-assessment of knowledge for those who tended not to provide 100% confidence judgements. This study provides evidence to the contrary, since the younger group, which made more 100% confidence judgements, displayed less overconfidence than the older participants.

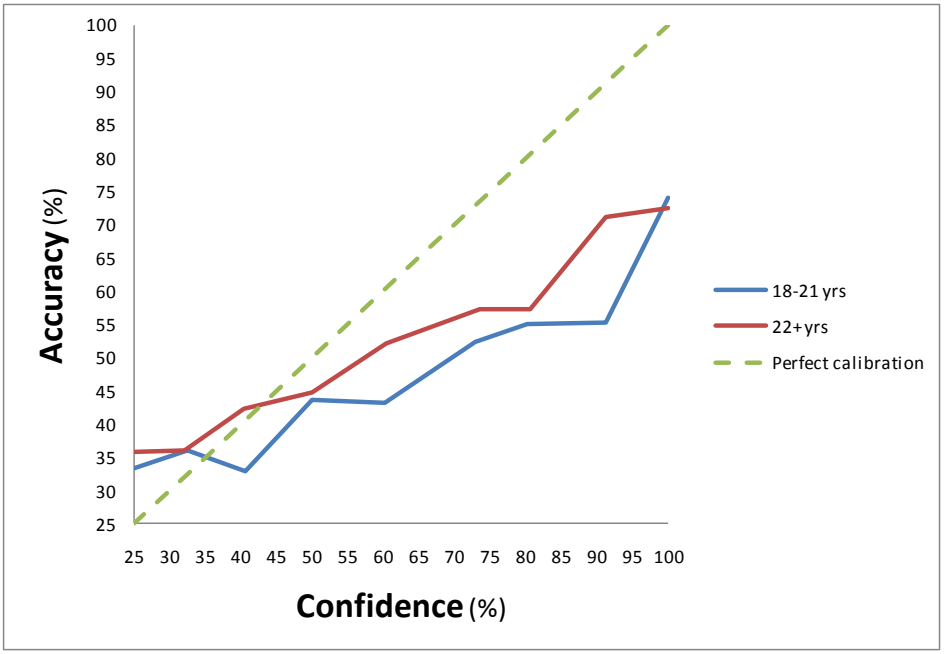


Figure 23: Calibration Curves by Age Group

4.4.3.2 Gender

The following hypothesis was tested in respect of gender:

H_0 = There is no gender difference in overconfidence

H_3 = Overconfidence is greater for male students

The results shown in Table 14 indicate that, while the positive mean bias score for each group indicated overconfidence, it was evident to a greater extent for males, with a mean bias score of 9.9%, as compared with 7.6% for females.

Table 14 – Confidence, Knowledge and Bias Score by Gender

	Mean Confidence %	*Knowledge Score %*	*Bias Score %*
Male	61.2	51.3	9.9
Female	56.3	48.7	7.6
Total	59.0	50.1	8.9

A t-test (see Appendix 17), indicated that this difference was statistically significant at the 5% level ($p=0.045<0.05$). Consequently, the null hypothesis is rejected and the results indicate greater overconfidence for males.

The calibration curves in Figure 24 show that results are similar by gender for judgements made with relatively high degrees of confidence. However at lower levels, females were more accurate and tended to display underconfidence for judgements made with confidence levels below 50%. For high confidence judgements, accuracy for questions in which 100% confidence was expressed was the same for each group at 73%. However, males were more inclined to express this degree of confidence, doing so for 24% of questions, as compared with 21% for females.

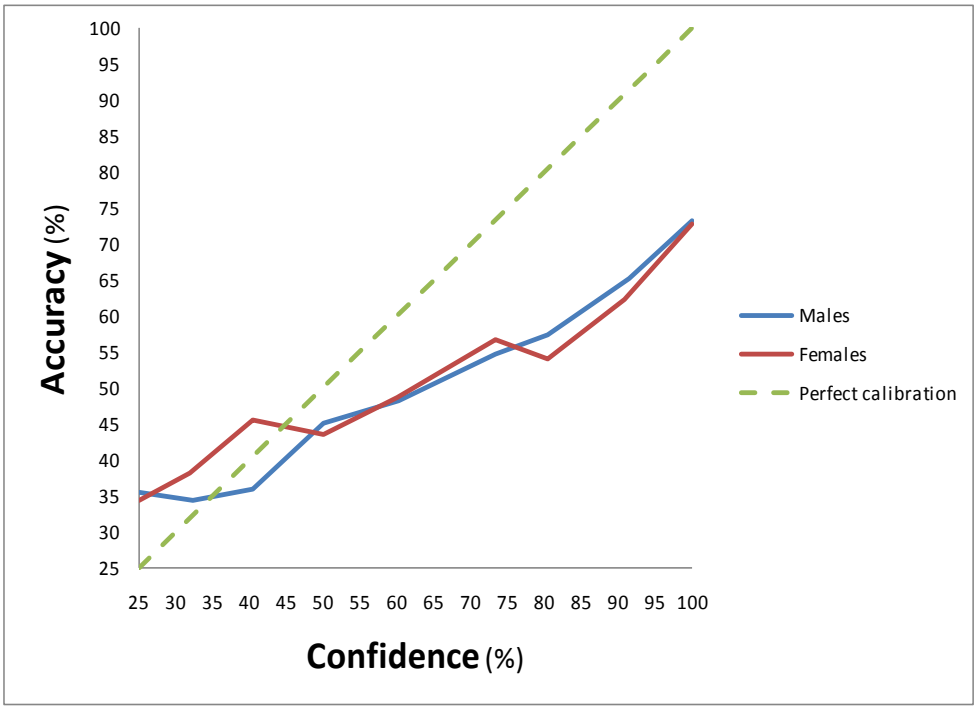


Figure 24: Calibration Curves by Gender

These findings support others reporting a greater tendency for overconfidence in males (Acker and Duck, 2008; Pallier, 2003; Barber and Odean, 2001; Lundeberg, Fox, and Punccohar, 1994). However, since as a group, females also displayed overconfidence, albeit to a lesser extent than males, they do not support Beyer's (1990 p.960) argument,

that 'self-derogatory' bias tends to result in females underestimating their ability. Rather, it supports the view that gender differences result, not from a lack of confidence in females, but rather a tendency for its excess in males (Lundeberg, Fox, and Punccohar, 1994).

Gender and Age

The potential moderating effect of age on gender was also explored and as shown in Table 15, bias score was higher for males regardless of their age group. A two factor ANOVA test (see Appendix 18) showed that there was no significant interaction between gender and age ($p=0.213>0.05$) and therefore, that gender differences are not moderated by age group.

Table 15 – Bias Score by Gender and Age

	18-21 Years %	*22+ Years* %	*Total* %
Male	7.8	13.3	10.0
Female	6.7	8.9	7.6
Total	7.3	11.3	8.9

4.4.3.3 Country of Origin

The following hypothesis was tested in respect of participants' country of origin:

H_0 = There is no difference in overconfidence between Chinese and UK students

H_4 = Overconfidence is greater for Chinese than UK students

Table 16 shows that, while both Chinese and UK students demonstrated overconfidence, with a positive mean bias score in each case, it was higher for Chinese students. The t-test shown in Appendix 19 indicates that this difference is statistically

significant at the 1% level ($p=0.001<0.01$). Consequently, the null hypothesis is rejected and the results indicate greater overconfidence for Chinese than UK students.

Table 16 – Confidence, Knowledge and Bias Score for UK and Chinese Students

	Mean Confidence %	*Knowledge Score %*	*Bias Score %*
UK	57.2	52.1	5.1
China	62.5	46.3	16.2

The calibration curve in Figure 25 shows different patterns of overconfidence, with Chinese students particularly overconfident for judgements made with high levels of confidence. They also showed a much greater tendency to express extremely high levels of confidence, with 30% of questions being answered with 100% confidence as compared with 20% for UK students. As indicated on the calibration chart, these judgements also resulted in a lower accuracy rate for Chinese participants, with 67% of such questions being answered correctly, as compared with 75% for UK students.

The higher levels of overconfidence detected for Chinese participants, as compared with UK students, supports previously reported higher levels in Asians as compared with Westerners (Acker and Duck, 2008; Yates, Lee and Bush, 1997) and more specifically, higher levels for Chinese (Culpepper, Zhao and Lowery, 2002; Yates, Lee and Shinotsuka, 1996). Culpepper, Zhao and Lowery (2002) also suggest that this is particularly likely to be the case in respect of knowledge about specific facts, in respect of which a tendency for less debate between opposing views in Chinese culture may result in a tendency to use extreme responses. The findings here provide supportive evidence for this, since Chinese participants displayed a greater tendency to express a confidence level of 100%. The greater overconfidence shown by Chinese students was explored further to test for possible moderating effects of age and gender.

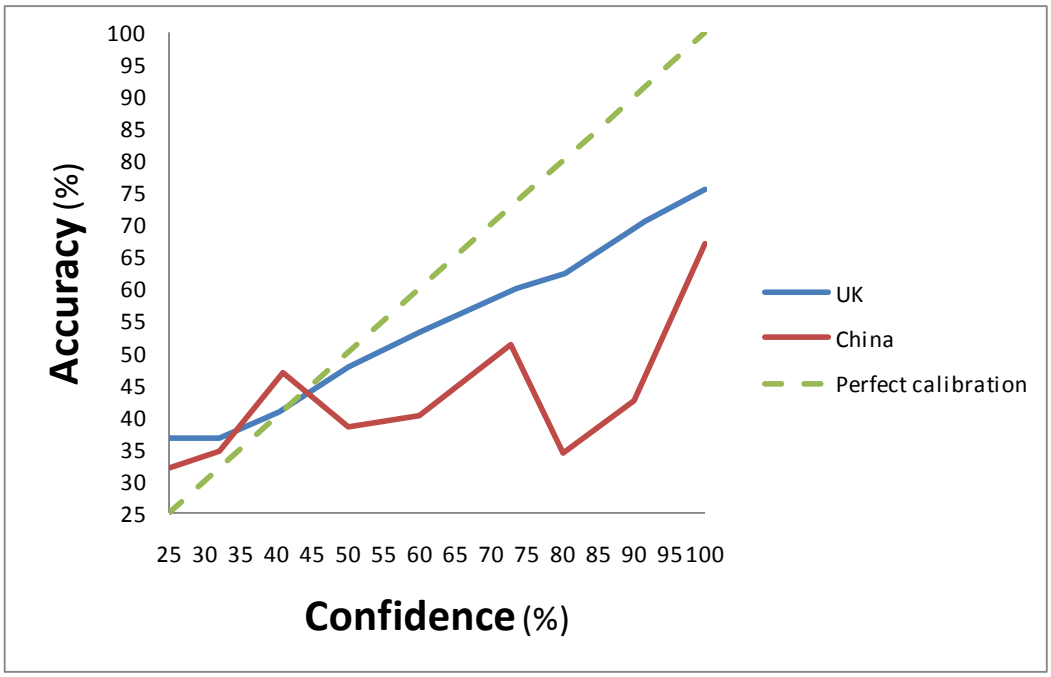


Figure 25: Calibration Curves by Country of Origin

Country of Origin and Age

Table 17 shows that the bias score is greater for Chinese students regardless of their age and also that, when analysed by country, the younger age group for both UK and Chinese participants, was more overconfident. A two factor ANOVA test (See Appendix 20) indicated no significant interaction between age and country of origin ($p=0.307>0.05$) and it can therefore be concluded that there is no evidence to support differences by country being moderated by age.

Table 17 – Bias Score by Country and Age Group

	18-21 Years %	*22+ Years* %	*Total* %
UK	5.2	4.3	5.1
China	20.0	14.9	16.2
Total	6.7	11.6	8.1

Country of Origin and Gender

Table 18 indicates that Chinese students were more overconfident, regardless of their gender. However, while in the case of UK students, males were more overconfident, for Chinese students this tendency was reversed. A two factor ANOVA (see Appendix 21) indicated a significant interaction between country of origin and gender at the 5% level ($p=0.040<0.05$) and therefore, that differences in overconfidence by gender were moderated by country of origin. It can therefore be concluded that, while there was a statistically significant finding that males were more overconfident than females, this was not so for Chinese students. While, as discussed earlier, investigation of all participants demonstrated statistically significant greater overconfidence for males, a t-test (See Appendix 22) indicated that for Chinese students there was no significant difference by gender ($p=0.535>0.05$). This supports Acker and Duck's (2008) finding of no significant difference between Asian males and females in terms overconfidence.

Table 18 – Bias Score by Country and Gender

	Male *%*	*Female* *%*	*Total* *%*
UK	7.0	2.3	5.1
China	15.1	17.0	16.2
Total	8.7	7.5	8.2

4.4.4 Overconfidence by Level of Study

Since participants comprised students studying at different levels at the university, differences by level were also investigated, including the potential moderating effects of age group, gender and country of origin. Table 19 shows that the mean bias score for each level was positive, indicating overconfidence in each case. However, the score for postgraduate students was higher than for each of the undergraduate groups, indicating greater overconfidence at that level. An ANOVA test (see Appendix 23) indicated that differences in the means were significant at the 1% level ($p=0.001<0.01$) and post hoc tests showed that these related to differences between level 7 and each of the other three levels.

Table 19 – Confidence, Knowledge and Bias Score by Level of Study

		Mean Confidence %	*Knowledge Score %*	*Bias Score %*
Undergraduates	*Level 4*	59.7	50.3	9.3
	Level 5	55.3	49.4	5.8
	Level 6	58.2	52.0	6.2
Postgraduates	*Level 7*	64.4	50.3	14.1
	Total	59.1	50.2	8.9

Consequently, in respect of undergraduate students, these results provide no evidence to challenge Russo and Schoemaker's (1992) view that metaknowledge tends not to be developed during formal education, since there is no statistically significant reduction in the bias score across these levels. Nor does it support the view that epistemological beliefs are an important influence on cognitive monitoring (Tynjala, Helle and Murtonen, 2001), when taking into account Jehng, Johnson and Anderson's (1993) findings that those studying at postgraduate level are more likely than undergraduates to recognise that knowledge is uncertain. In this case, postgraduates displayed greater overconfidence than undergraduates. To explore these findings further, overconfidence by level of study was tested for potential moderating effects of individual differences in respect of age, gender or country of origin.

4.4.4.1 Level of Study and Age

Table 20 shows age differences in the bias score by level of study and indicates that postgraduates were more overconfident than each undergraduate group, regardless of their age group. A two factor ANOVA test (See Appendix 24) confirmed that there was no significant interaction between level of study and age group ($p=0.856>0.05$) and consequently, that study level differences are not moderated by age group.

Table 20 – Bias Score by Level of Study and Age

		18-21 years *%*	*22+ years* *%*	*Total* *%*
Undergraduates	*Level 4*	9.2	10.4	9.3
	Level 5	5.5	8.3	5.8
	Level 6	7.7	6.2	6.6
Postgraduates	*Level 7*	14.4	14.4	14.2
	Total	7.3	11.3	8.9

4.4.4.2 Level of Study and Gender

Table 21 shows that postgraduate students were more overconfident than any of the undergraduate groups, regardless of their gender and a two way ANOVA test (see Appendix 25) confirmed that there was no significant interaction between these variables ($p=0.218>0.05$). Consequently, there is no evidence that study level differences are moderated by gender.

Table 21 – Bias Score by Level of Study and Gender

		Male *%*	*Female* *%*	*Total* *%*
Undergraduates	*Level 4*	10.8	7.5	9.4
	Level 5	5.5	6.3	5.8
	Level 6	6.7	5.8	6.2
Postgraduates	*Level 7*	16.8	10.9	14.2
	Total	9.9	7.6	8.9

4.4.4.3 Level of Study and Country of Origin

The majority of full time postgraduate business students at the institution in which the study was conducted are non-UK students and during the period during which the data was collected, 42% of the population were Chinese and 9% UK students. This was

reflected in the sample, in which a significant proportion (39%) of level 7 students was Chinese, and only a small minority (8%), UK students. Consequently, given the higher levels of overconfidence displayed by Chinese participants, the potential moderating effect of this on differences between study levels was explored. Figure 26 shows a comparison of bias scores for UK and Chinese students by level of study and country of origin. A two factor ANOVA test (See Appendix 26) indicated that interactions between study level and country were significant at the 1% level ($p=0.009<0.01$). However, as shown in Figure 26, while greater for other study levels, differences in overconfidence by country for postgraduates (level 7), were relatively small, with a mean bias score for UK students of 15% as compared with 16.4% for Chinese (See Appendix 26). Consequently, there is no evidence that high levels of overconfidence for postgraduate students are moderated by country of origin.

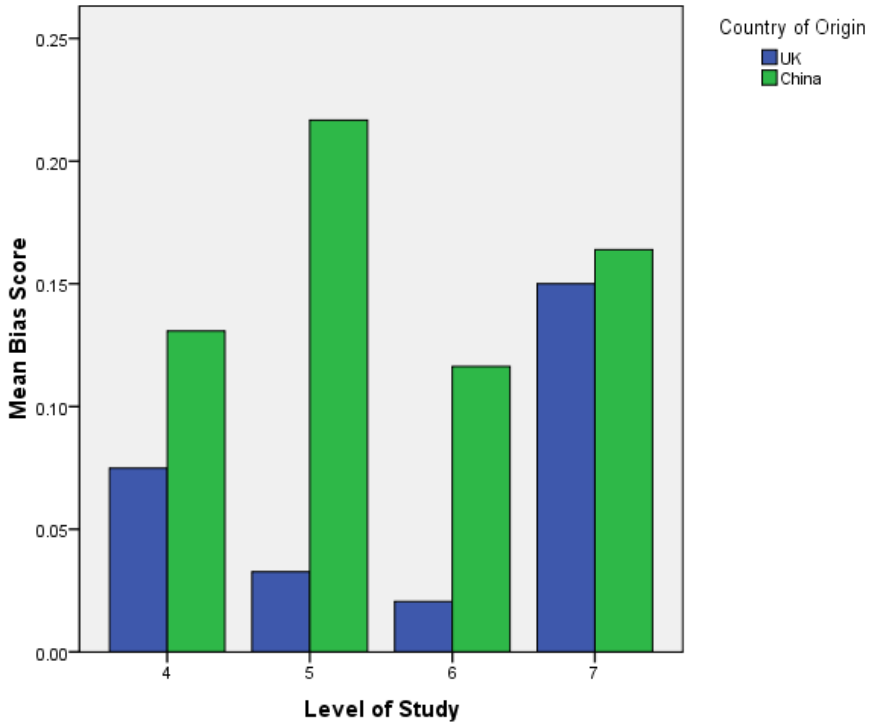


Figure 26: Mean Bias Scores by Level of Study and Country of Origin

4.4.4.4 Entry Status

Students may enter the university either as new students (direct entrants) or as continuing students, progressing in their studies from a previous level of study. For example, the university admits students with appropriate qualifications directly to the final year of undergraduate programmes. Consequently, those studying at level 6 can be grouped according to whether they entered directly at that stage or progressed from level 5. Similarly, postgraduate students may enter the university directly onto a postgraduate programme (level 7), or progress internally from a previous undergraduate programme at level 6. Therefore, differences in overconfidence bias between direct entrants and continuing students at each of these levels were investigated. Table 22 shows that at level 6, the mean bias score for direct entrants exceeds that of continuing students and a t-test (See Appendix 27) indicated that this difference was significant at the 5% level ($p=0.042<0.05$). Consequently, it can be concluded that those entering the university directly onto the final stage of undergraduate programmes were more overconfident than continuing undergraduate students, who had previous experience of studying at the institution. For postgraduate students, while both groups were overconfident, this was evident to a greater extent for continuing students. However, a t-test (See Appendix 28) indicated that this difference was not statistically significant ($p=0.065>0.05$), although it should be noted that the p-value is close to the 5% level.

Table 22 – Confidence, Knowledge and Bias Score by Entry Status

		Mean Confidence %	*Knowledge Score %*	*Bias Score %*
Undergraduate Level 6	Direct entrants	60.9	52.0	8.9
	Continuing students	54.3	52.9	2.4
Postgraduate Level 7	Direct entrants	63.5	51.3	12.2
	Continuing students	65.9	48.4	17.5

4.4.5 Overconfidence and Academic Performance

The association between overconfidence and academic achievement on the participants' study programmes was investigated by testing the following hypothesis:

H_0 = There is a no association between overconfidence and academic performance

H_5 = There is a negative association between overconfidence and academic performance

Data in respect of academic performance was unavailable for 39 participants (8%), due to them not supplying their student registration number, supplying an incorrect number or terminating their studies before submitting work for assessment. For the remainder, Appendix 29 shows correlations between bias scores and marks obtained by participants for the stage of their study programme during which the confidence data was collected. These include the overall mark for that stage and, since these were achieved through a variety of assessment tasks, marks attained for different modes of assessment. The results show very strong positive correlation, significant at the 1% level ($p=0.001<0.01$), between overall assessment marks and those for examinations ($r=+0.867$), coursework ($r=+0.855$) and dissertations ($r=+0.739$). This indicates that higher achieving students tended to perform consistently better across these different modes of assessment and poorer performers, worse.

4.4.5.1 Overall Academic Performance

Investigation of the association between overconfidence and overall academic achievement revealed weak and negative, but, due to the large sample size, statistically significant at the 1% level, correlation between bias score and students' overall marks ($r=-0.113$, $p=0.007<0.01$). This indicates that the more overconfident in their knowledge participants were, the lower their overall level of academic achievement was. Consequently, the null hypothesis is rejected and the results indicate that greater overconfidence is associated with poorer academic performance. This finding is in accordance with the view that less competent individuals tend to experience greater difficulty when making metacognitive judgements than the more competent (Dunning et

al, 2003) and more specifically, literature highlighting an inverse relationship between overconfidence and academic competence (Pieschl, 2009; Ehrlinger et al, 2008; Paris and Paris, 2001; Everson and Tobias, 1998). It also supports Koku and Qureshi's (2004) findings in a smaller scale investigation of business students. However, while in that research, the same instrument was used to assess both overconfidence and academic performance, in this study the latter was determined independently, using student marks achieved during their study programme. While the finding here indicates association rather than causality, it does provides supportive evidence for Grimes' (2002) argument that the abilities required for academic achievement are associated with those necessary for accurate self-assessment. It also lends weight to the argument that for some, poor self-assessment accuracy can result in a dual burden, in which they possess not only poor knowledge, but also insufficiently developed metaknowledge with which to appreciate this (Kruger and Dunning, 1999). However, as noted above, negative correlation between academic performance and overconfidence, while statistically significant, is weak. Therefore to investigate this issue further, participants were divided into three groups of equal size according to their academic ability, as indicated by their overall assessment mark on their study programme. High performers were those achieving at least 63.1%, the low performing group achieved below 54.3% and the remainder were assigned to the middle group. As shown in Table 23, those in the lower performing group were most overconfident in their knowledge, with a mean bias score of 10.5%, followed by the middle performing group (9.6%). The higher performers were least overconfident, with a mean bias score of 7%. However, an ANOVA test (See Appendix 30) indicated that these differences were not significant at the 5% level ($p=0.099>0.05$).

Table 23 – Mean Bias Score by Academic Performance Group

	Mean Bias Score %
High performing	7.0
Middle performing	9.6
Low performing	10.5

Differences between performance groups in this case are less than those detected in Koku and Qureshi's (2004) study of business students. In their case, they reported bias scores for three similarly compiled academic performance groups, of 3% for high performers, 10% for the middle group and 22% for low performers. However academic performance was more narrowly defined in their study, in that it was determined using only scores achieved in a multiple choice examination while, in this study, a broader indicator of performance was used, incorporating all marks achieved across each participant's study programme. As noted earlier, this also entailed using an indicator of academic performance which did not rely on the same data used to determine overconfidence. The findings here therefore provide evidence that when using this approach, differences in overconfidence between achievement groups, are less than those detected by Koku and Qureshi (2004).

4.4.5.2 Performance in Different Modes of Assessment

As well as overall academic performance, Appendix 29 also shows correlations between the bias score and academic performance in different modes of assessment. These indicate weak, but statistically significant, negative correlation in respect of examinations ($r=-0.096$, $p=0.019<0.05$) and coursework ($r=-0.138$, $p=0.001<0.01$). These correlations are similar to that between the bias score and overall academic performance ($r=-0.113$) due to the high positive correlation between the latter and achievement in examinations and coursework and indicate that overconfidence is only weakly associated with poorer performance in these types of assessment. Performance in the other two modes of assessment, dissertations ($r=-0.004$, $p=0.479>0.05$) and oral presentations ($r=0.045$, $p=0.348>0.05$) show statistically insignificant and very weak, correlation with the bias score. For dissertations, a possible reason for this stems from the fact that, unlike other study modules, students receive individual supervision from an academic member of staff when undertaking this work. Individual and focused formative feedback received during this process may therefore help to reduce the impact of overconfidence in knowledge on work which is subsequently submitted for assessment. In the case of presentations, while, as discussed earlier, much emphasis is placed on the potential negative consequences of overconfidence, possible advantages have also been highlighted. Since confidence can be associated with competence,

especially in circumstances where it is difficult to assess (Klayman et al, 1999), unwarranted displays of confidence may attract favourable attention and rewards (Fischhoff 1994). Consequently, it may be that overconfident students attract this attention and are rewarded for their confidence, despite the fact that it may be misplaced, when delivering presentations. Another issue which may impact on the association between overconfidence in knowledge and academic achievement is the extent to which knowledge is rewarded in the assessment process. Pieschl (2009) argues that students' ability to accurately assess their own knowledge is important for more complex learning tasks, as well as lower level activities. However, in higher education knowledge acquisition tends to be rewarded more highly at lower levels of study, while at higher levels there is more emphasis on more advanced learning outcomes. Since more demanding skills are required as students progress through different levels of study, Isaacson and Fujita (2006) highlight the potential value in exploring whether effective knowledge monitoring becomes increasingly important for academic tasks that require higher level learning skills. Consequently, the association between overconfidence and academic performance at different levels of study was investigated.

4.4.5.3 Overconfidence and Academic Performance by Level of Study

Correlations between bias score and academic performance for each study level are shown in Appendices 31-34. Table 24 summarises these findings in respect of overall academic performance and indicates that while negative correlation is weak in each case, it is statistically significant at the 5% level, for both undergraduates studying at level 4 and postgraduate students and each of these are discussed below.

Table 24 – Correlation between Bias Score and Overall Marks by Level of Study

		R	*p (one tailed)*
Undergraduates	*Level 4*	-0.188*	0.021
	Level 5	-0.055	0.247
	Level 6	-0.071	0.276
Postgraduates	*Level 7*	-0.175*	0.026

**Significant at 5% level (one tailed)*

Undergraduate Level 4 Students

The stronger negative association between overconfidence and academic performance for undergraduates studying at level 4, as compared with levels 5 and 6, may be attributable to knowledge being more highly rewarded at level 4 than at higher levels of study, where there is greater emphasis on more advanced learning outcomes. While it may usefully inform learning in other situations, effective self-monitoring of knowledge may be more influential in situations in which the ability to demonstrate knowledge is critical.

Alternatively, since these students are new to both the institution, and higher education, they may not yet have developed other metacognitive knowledge which may usefully inform metacognitive regulatory activity in that learning environment to an extent that it compensates for their poor self-monitoring ability. For example, while undergraduates who have more experience of studying in higher education may also be overconfident about how much they know, they may have developed greater declarative task knowledge of learning strategies, procedural knowledge of how to use them and conditional knowledge of when they are most appropriate in a higher education learning environment. This enhanced metacognitive knowledge may positively impact on their learning and academic performance to an extent that offsets the adverse consequences of poor self-monitoring.

In terms of the association between the bias score and each mode of assessment for level 4 students, the results shown in Appendix 31 indicate weak, but statistically significant at the 5% level, negative correlation for examinations ($r=-0.156$, $p=0.047<0.05$) and coursework ($r=-0.212$, $p=0.011<0.05$). This indicates that while overconfidence is associated with poor academic performance in each of these to some extent, like overall academic performance, this association is not strong. For assessment based on oral presentations, correlation is positive but not significant ($r=0.092$, $p=0.280>0.05$) and as discussed above, the lack of association between overconfidence and performance in this form of assessment may be attributable to being rewarded for confidence when delivering presentations.

Postgraduate Students

Given the higher level learning outcomes associated with postgraduate study, the implications of greater emphasis on knowledge in assessment advanced for level 4 students above is clearly inappropriate in the case of postgraduate students. Investigation of the association between the bias score and performance in each mode of assessment for this group also indicates weakly negative correlation in respect of all modes of assessment except presentations (see Appendix 34). However, for postgraduates, this correlation is not statistically significant at the 5% level for any of these.

Differences according to entry status were also investigated, by separately analysing direct entrants and continuing students (See Appendices 35 and 36). The results show that for direct entrants, there is a weak, but statistically significant, negative association between the bias score and both overall performance ($r=-0.226$, $p=0.022<0.05$), examinations ($r=-0.192$, $p=0.049<0.05$) and coursework marks ($r=-0.241$, $p=0.016<0.05$). However, for continuing students, there was no significant association between overconfidence and any aspect of academic performance. Whilst these associations do not necessarily indicate a causal relationship, they suggest that the overconfidence of students new to the institution for their postgraduate studies may have more influence on their academic performance than those who have progressed from undergraduate studies and have been with the university for some time and are therefore familiar with the specific demands of the learning environment. To investigate this issue further and in a broader context, direct entrants were compared with continuing students for all levels of study. While as indicated in Table 25, direct entrants were more overconfident, with a higher mean bias score than continuing students, a t-test (see Appendix 37) indicated the difference between the two groups was not statistically significantly at the 5% level ($p=0.092>0.05$)

Table 25

Mean Confidence, Knowledge and Bias Score for All Students by Entry Status

	Mean Confidence %	*Knowledge Score %*	*Bias Score %*
Direct entrants	60.8	50.8	10.0
Continuing students	57.4	49.6	7.8

However, while the results did not reveal a statistically significant difference in levels of overconfidence between each group, there were differences between each group in terms of its association with academic performance. For direct entrants, Appendix 38 shows weak and negative, but statistically significant at the 1% level, correlation between bias score and overall academic performance ($r=-0.190$, $p=0.002<0.01$) as well as examination ($r=-0.171$, $p=0.004<0.01$) and coursework marks ($r=-0.199$, $p=0.001<0.01$). However, for continuing students there was no statistically significant correlation between bias score and any aspect of academic performance (see Appendix 39). Consequently, these findings indicate that, while differences in bias scores between those new to the university and continuing students are not statistically significant, overconfidence may have a greater impact on academic performance for new students. As suggested earlier, a possible explanation for this is that continuing students, who have experience of the demands of higher education generally, and their study programme specifically, have developed better declarative task, procedural and conditional metacognitive knowledge. Since this would equip them better to meet the assessment demands of their study programmes, it may compensate for poor metaknowledge.

4.4.5.4 Overconfidence and Academic Performance – Individual Differences

Since the association between overconfidence in knowledge and academic performance may be subject to individual differences, this was investigated in respect of age, gender and country of origin.

Age

Appendices 40 and 41 show results for the investigation of the association between the bias score and academic performance for each age group. These show that for the 18-21 years group, correlation is not significant in respect of the overall assessment mark, or any individual mode of assessment ($p>0.05$). However, for older students, there is weak, negative correlation between bias score and coursework, which is significant at the 5% level ($r=-0.153$, $p= 0.016<0.05$). This indicates that for this group, greater overconfidence is associated with poorer performance in coursework, indicating that not only are they more overconfident about their knowledge than the younger participants, but that this is also linked more closely with academic performance in this mode of assessment, albeit not strongly.

Gender

Investigation of the association between bias score and academic performance for males and females separately, indicate only small gender differences (See Appendices 42 and 43). For males, the results reveal weak, but statistically significant, negative correlation between bias score and each of the overall mark ($r=-0.117$, $p=0.030<0.05$) and coursework marks ($r=-0.120$, $p=0.026<0.05$). Correlation with performance in examinations is also close to significant at the 5% level ($r=-.099$, $p=0.055$). For females, the findings are similar. Weak correlation between bias score and overall marks is close to significant at the 5% level ($r=-0.113$, $p=0.054>0.05$) and for coursework, weak negative correlation is significant at the 1% level ($r=-0.171$, $p=0.007<0.01$).

Country of Origin

Appendices 44 and 45 show results for the investigation of associations between bias score and academic performance for UK and Chinese students. For UK participants, they indicate no statistically significant association between bias score and overall academic performance, or achievement in any of the different modes of assessment ($p>0.05$). However, for Chinese students, there is moderately strong negative correlation between bias score and marks for oral presentations, which is statistically significant at the 5% level ($r=-0.595$, $p=0.035<0.05$). This finding is interesting given that this association has, unlike that between bias score and other modes of assessment, tended to be positive when investigated earlier in other dimensions. It indicates that for Chinese students, unlike other groups, greater overconfidence is associated with worse performance in assessed presentations. This suggests that if for other students, overconfidence in knowledge may translate into confident presentations, which are rewarded in the assessment process; this is not the case for Chinese students who are also confident about their knowledge. While an explanation for this is beyond the scope of this study, it may relate to the theory discussed earlier suggesting that the behaviour of this group may tend to differ according to the nature of the task. It has been argued that nomothetic enquiries based on specific facts, such as that used to collect data for this study, may increase the potential for overconfidence in Chinese respondents, due to their tendency to consider less contrary evidence (Culpepper, Zhao and Lowery, 2002). However, it has also been suggested that there is a greater tendency for modesty and self-effacement in more idiographic tasks in Chinese culture (Yates, Lee and Shinotsuka, 1996). Consequently, it may be that when comparing those who are overconfident in their knowledge, Chinese students are less likely to display the type of confident behaviour that other students are rewarded for when delivering assessed presentations.

4.5 Discussion and Implications for Professional Practice

Table 26 summarises the findings of the study in respect of the research hypotheses established in Chapter 2 and these are discussed below.

Table 26 - Summary of Results from Hypothesis Testing

Hypothesis	***Outcome***	***Significance Level***
H_1: Students are overconfident in their knowledge	Accepted	1%
H_2: Overconfidence differs between older and younger students	Accepted *(Older students are more overconfident)*	1%
H_3: Overconfidence is greater for male students	Accepted	5%
H_4: Overconfidence is greater for Chinese than UK students	Accepted	1%
H_5: There is a negative association between overconfidence and academic performance	Accepted *(Weak association)*	1%

4.5.1 Overconfidence

The mean bias score across all participants was 8.9%, indicating a general tendency for overconfidence in knowledge, which was confirmed as being statistically significant. Positive correlation between confidence and overconfidence also indicated that those who displayed more confidence also tended to demonstrate greater overconfidence. The results also indicated a statistically significant negative correlation between knowledge and bias scores, indicating that those who were most overconfident tended to know less about the subject matter on which they were tested.

4.5.2 Individual Differences

Table 27 summarises findings from the investigation of individual differences in respect of age, gender and country of origin, as well as other sub-groupings investigated.

Table 27 – Summary of Individual Differences in Bias Score

		Bias Score *(%)*	**Sig.**
All Participants		8.9	✓✓
Age Group	18-21 yrs	7.3	✓✓
	22+ yrs	11.3	
Gender	Male	9.9	✓
	Female	7.6	
Country of Origin	UK	5.1	✓✓
	China	16.2	
Level of Study	Undergrad Level 4	9.3	✓✓
	Undergrad Level 5	5.8	
	Undergrad Level 6	6.2	
	Postgraduate	14.1	
Entry Status	Direct entrants	10.0	X
	Continuing students	7.8	
Significance levels in group differences: ✓✓ ***1%*** ✓ ***5%*** X *not significant*			

Due to the nature of the population, analysis of the association between age and overconfidence was restricted to some extent, since the age distribution was relatively narrow. However, the results indicated that those aged 22 years and above displayed higher levels of overconfidence than those aged 18-21 years. This finding suggests that interventions to assist in moderating overconfidence may be of greater value in learning environments and study programmes with a higher proportion of older students.

In respect of gender, while both groups displayed overconfidence, males displayed statistically significant higher levels than females. This suggests that initiatives to enhance metaknowledge may be more important for programmes with a higher proportion of male participants. However, in many business programmes both gender groups tend to be relatively well represented and therefore, this finding may be most usefully employed in highlighting the gender bias to learners in order that males in particular, who are more likely to display overconfidence, may consider the potential implications of this and how it may be moderated.

With regard to country of origin, Chinese were more overconfident than UK students. There are business study programmes offered by the institution in question which have a relatively high proportion of Chinese students and this finding suggests that interventions to enhance metaknowledge may be particularly appropriate in these. Alternatively, induction programmes which are designed by the institution specifically for overseas students could include initiatives to highlight this issue and assist in the development of metaknowledge. Further analysis also revealed a statistically significant interaction between country of origin and gender, with females displaying greater overconfidence in the case of Chinese and males for UK students and this is another issue which could be brought to the attention of Chinese students.

4.5.2.1 Level of Study

Investigation of differences by level of study indicated overconfidence for undergraduates at all three levels, but that differences between them were not statistically significant. This is not in accordance with findings in Kennedy, Lawton and Plumlee's (2002) study of sociology and business students, in which overconfidence declined with the length of time participants had spent on their courses and it was suggested that the educational experience had helped to indicate limitations in their knowledge. Instead, it supports the view that metaknowledge tends not to be developed during formal education (Russo and Schoemaker, 1992) and suggests the need for greater emphasis on its development, to enable students to better appreciate gaps in their knowledge. This would enable them to adapt their learning strategies accordingly since,

if an overconfident individual believes their knowledge is good, they will not prompted to improve it (Yates, Lee and Shinotsuka, 1996)

Interestingly, in view of Jehng, Johnson and Anderson's (1993) findings, indicating that postgraduate students have a lower tendency to see knowledge as certain, in this study they displayed statistically significant greater overconfidence than those studying at all three levels of undergraduate study. Investigation of the potential moderating effect of demographic differences on this finding indicated no statistically significant interactions in respect of age or gender. In the case of country of origin, while there was a significant interaction, this did not relate to postgraduates, for whom levels of overconfidence were similar between UK and Chinese students. Consequently, initiatives to assist in the development of metaknowledge may be particularly beneficial for those studying at postgraduate level. However, while this may assist in developing their self-monitoring ability, the potential for this greater capability to impact positively on their academic performance in higher education will be influenced by the association between overconfidence in knowledge and academic achievement.

4.5.3 Overconfidence and Academic Performance

If learners are overconfident in their knowledge, this may adversely affect self regulated learning, as it could influence the development of learning strategies in response to self-monitoring activity (Stone, 2000). Ehrlinger et al (2008) argue that we should be concerned about poorly performing students whose overconfidence may prevent them from improving through additional study. These arguments raise the question of whether there is an association between overconfidence and poor academic performance. Findings in respect of this study are summarised in Table 28.

Table 28 – Correlation between Bias Score and Academic Performance

		Correlation Between Bias Score and Academic Performance									
		Overall Performance		Exams		Presentations		Coursework		Dissertations	
		r	Sig	r	Sig.	r	Sig.	r	Sig.	r	Sig.
All Participants		**-.113**	**√√**	**-.096**	**√**	.045	X	**-.138**	**√√**	-.004	X
Level of Study	Undergrad. Level 4	**-.188**	**√**	**-.156**	**√**	.092	X	**-.212**	**√**	N/A	N/A
	Undergrad. Level 5	-.055	X	-.051	X	-.147	X	-.111	X	N/A	N/A
	Undergrad. Level 6	-.071	X	-.043	X	-.425	X	-.129	X	-.008	X
	Postgrad.	**-.175**	**√**	-.122	X	.009	X	-.126	X	-.029	X
Entry Status	Direct entrants	**-.190**	**√√**	**-.171**	**√√**	.022	X	**-.199**	**√√**	-.077	X
	Continuing students	-.023	X	-.021	X	-.206	X	-.067	X	.077	X
Age Group	18-21 yrs	-.074	X	-.061	X	.061	X	-.070	X	-.219	X
	22+ yrs	-.110	X	-.078	X	.059	X	**-.153**	**√**	.001	X
Gender	Male	**-.117**	**√**	-.099	X	.068	X	**-.120**	**√**	.036	X
	Female	-.113	X	-.101	X	.042	X	**-.171**	**√√**	-.058	X
Country of Origin	UK	.013	X	.030	X	.126	X	.048	X	.119	X
	China	-.008	X	-.082	X	**-.595**	**√**	-.010	X	.167	X
Significance levels: **√√** ***1%*** **√** ***5%*** X *not significant*											

The study revealed weak, but statistically significant, negative correlation between bias score and overall academic achievement for the sample as a whole, which indicates that

greater overconfidence was to some extent associated with poorer levels of achievement. While causality has not been established, this finding suggests that developing better metaknowledge may be influential in enhancing learning and academic performance. However, since the association is weak it provides only limited supportive evidence that learning can be improved through better appreciation of gaps in understanding (Schraw, 1998), and that this is particularly the case in an environment, such as higher education, which requires large amounts of information to be learned (Clarebout, Elen and Onghena, 2006). This issue was explored further by investigating the association between overconfidence and academic performance in different methods used to assess student learning, to determine whether the association differed across methods. This indicated weak, but statistically significant, negative association between overconfidence and achievement in both examinations and coursework, which suggests that, while it may be possible to improve performance in each of these through the development of metaknowledge, the improvement may not be striking.

The association between overconfidence and academic performance was also investigated by level of study, to determine whether it differed accordingly. This indicated statistically significant, negative association for two groups, those at the first level of study on undergraduate programmes and postgraduates. The closer association between overconfidence and academic performance for first level, as compared with other undergraduates, may be due to the greater emphasis placed on demonstrating knowledge at this level of study. In such circumstances, those who overestimate their own knowledge may perform worse when their learning is assessed. However, this does not explain the finding for postgraduate students, for whom there is greater emphasis on higher level learning outcomes, such as analysis and critical evaluation. Consequently, to investigate this group in greater detail, they were analysed according to whether they had joined the institution directly for postgraduate study, or had progressed to that level from previous undergraduate studies at the university. This revealed statistically significant negative correlation between overconfidence and overall academic performance for direct entrants but not continuing students. Extending this analysis to all students participating in the study revealed similar results, indicating that, while the difference in overconfidence between each group was not statistically significant, its

association with academic performance was. This suggests that for students new to the university, overconfidence in knowledge may be more likely to adversely impact on their academic performance than those who have progressed from a previous level of study and are more familiar with the institution and its academic demands. Consequently, while the study suggests that initiatives to assist in the development of metaknowledge may have some, albeit limited, implications for academic performance, they may be better targeted at new entrants, perhaps during induction activities designed to help prepare them for their studies at the university, or during the early part of their study programme.

Individual differences in the association between overconfidence and academic performance were also investigated in respect of age, gender and country of origin. In the case of age, differences were not striking, with a statistically significant association indicated only for the older (22 years +) group in respect of weak negative correlation between overconfidence and coursework marks. Analysis by gender also revealed similar findings for each sub-group, with statistically significant negative associations between overconfidence and overall performance for males and close to significance at the 5% level for females (p=.054), as well as for coursework for each group. These findings suggest that overconfidence is more closely associated with academic performance when students are assessed by coursework than the other modes of assessment investigated; examinations and oral presentations. Table 28 shows that for the participants as a whole, as well as for each sub-group (with the exception of Chinese students) where negative correlation is statistically significant, it is stronger for coursework than any other type of assessment. This suggests that initiatives aimed at improving metaknowledge and reducing overconfidence may have greater impact on academic performance in study programmes placing a greater emphasis on coursework. In respect of country of origin, separate analysis of UK and Chinese students revealed no statistically significant evidence of an association between overconfidence and overall academic performance for either group. In the case of Chinese students, this is particularly interesting, since as a group they displayed relatively high levels of overconfidence, which do not appear to be linked to how they perform during their study programmes.

In conclusion, while the study has produced findings which support the argument that accurate self-monitoring is essential for effective learning (Pieschl, 2009) to some extent, the evidence is not strong. Although it revealed a statistically significant negative association between overconfidence and overall academic performance for the sample as a whole, as well as various sub-groups, correlations are typically weak. Classifying students according to overall academic performance, the results also show that while overconfidence tended to be greater for lower achievement groups, differences were not statistically significant.

Investigation of individual differences also indicated statistically significant association between academic performance and overconfidence for some groups but not others. Consequently, the extent to which accurate self-monitoring of knowledge influences learning may depend on who is doing the learning and in what context. For example, while an association is evident for those new to the university, it is not for continuing students, nor, despite their relatively high levels of overconfidence, for Chinese students. This suggests that for some, overconfidence may not be an obstacle to learning, or that the extent to which it is, is not largely influential in their capacity to perform well in summative assessment activities. Alternatively, it may be that any disadvantages associated with overconfidence which *do* impact on assessment outcomes are moderated, to some extent, by other compensating influences. A potential example of this was highlighted when discussing findings related to the association between overconfidence and marks achieved for assessed presentations. In this case, overconfidence in knowledge may adversely impact on the learning of some students, due to their inability to recognise, and therefore respond appropriately to, gaps in their knowledge. However, if for example, overconfidence in knowledge is positively associated with confidence in other contexts, including self-assured delivery of oral presentations, this could compensate for any disadvantages in knowledge acquisition. A lack of strong evidence of better academic performance for those who are less overconfident may also relate to the extent to which such self-monitoring skills inform their learning, since to be effective, the outcomes from self-evaluation should inform decisions students' make in respect of their learning strategies (Klenowski, 1995). In this case it may be that potentially misleading outcomes from poor self-monitoring are not particularly influential because students do not rely on them greatly when devising

learning strategies anyway. In terms of the conceptual framework developed in Chapter 2, as illustrated in Figure 27, this would suggest that while poorly developed metacognitive self-knowledge reduces monitoring accuracy, its impact on academic achievement may be relatively low due to the limited influence knowledge monitoring has on planning learning activities.

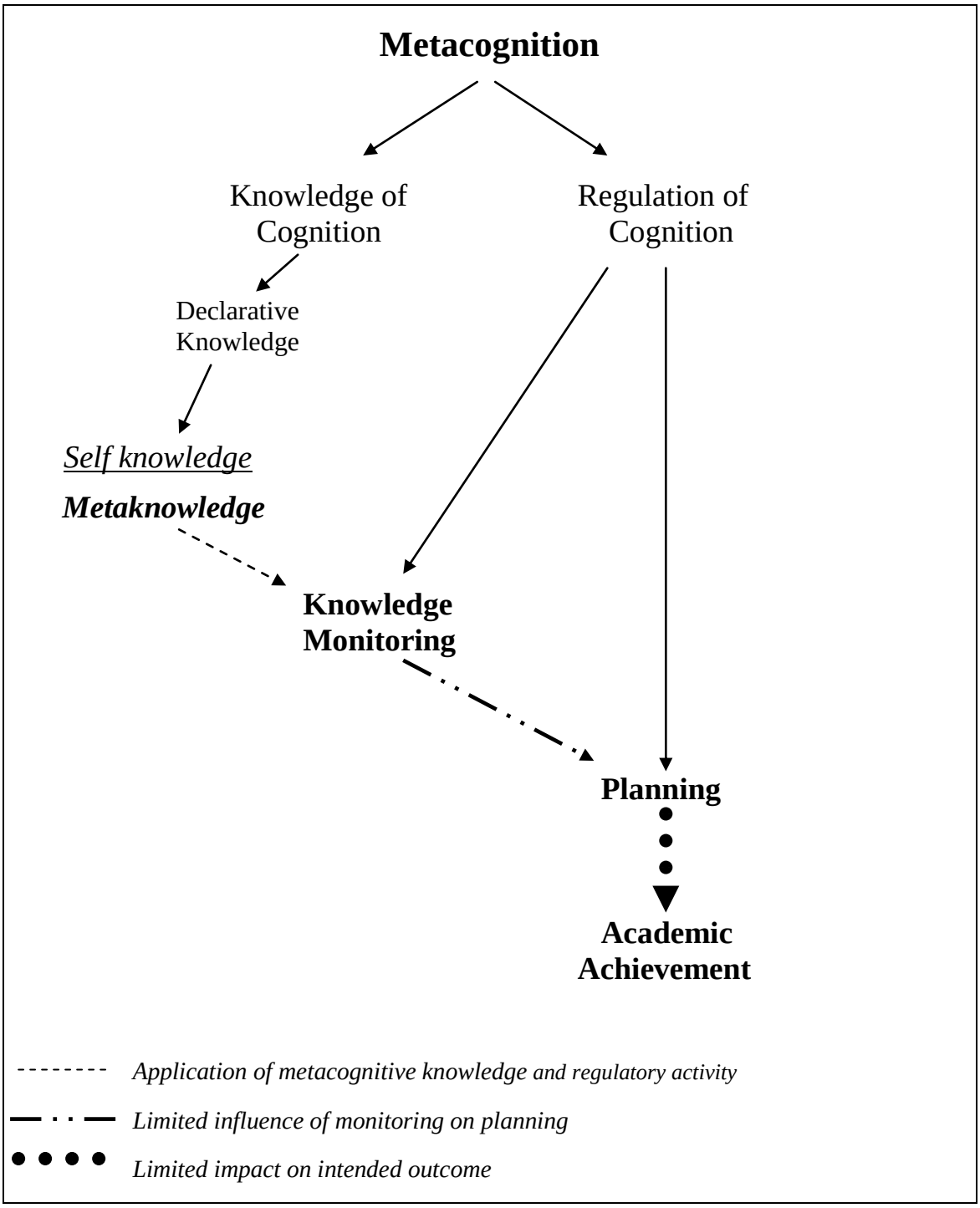


Figure 27: Limited Influence of Knowledge Monitoring on Planning

In terms of the implications of these findings for professional educators working in higher education, the absence of evidence of a strong association between accurate self-monitoring of knowledge and academic performance does not necessarily imply that there is no value in addressing metaknowledge in business programmes. Flavell (1979) argues that as well as learning more effectively, those with better metacognitive monitoring skills may be better equipped when taking lifestyle decisions. Additionally, Crittenden and Woodside (2007) argue that students and business managers should appreciate the importance of metathinking skills and that business schools should do better in developing students' appreciation of how these skills may improve their chances of success in business. Business management entails taking decisions in conditions of uncertainty and overconfidence in knowledge is therefore a potential impediment. Ramnarayan, Strohschneider and Schaub (1997), for example, in a study of advanced students in a school of management, analysed factors that contributed to poor performance in a business simulation and found that participants typically displayed proficiency in terms of basic knowledge, but lacked metaknowledge. Consequently, interventions to address the consequences of poor metaknowledge may usefully inform students in these respects and help to equip them for a career in business management, as well as highlighting its potential consequences for learning.

Having discussed the findings from the research in detail, the following chapter will draw conclusions in respect of the study and these findings and their contribution to the understanding of how effective learners are in monitoring their own knowledge and the implications of this.

5. CONCLUSIONS

5.1 Introduction

This chapter draws conclusions from the research and discusses its contribution towards answering central questions of how effectively individuals are able to self-assess their capabilities and what the implications of this are. Before doing so, it is important to acknowledge that attempts to explain the world are fallible, cannot be justified absolutely and are subject to critique and replacement by alternative explanations (Scott, 2005). Consequently, when reporting on research studies it is important not to be overconfident about conclusions, nor to over-generalise (Wallace and Wray, 2006). This principle is, of course, particularly apt in a study investigating the theory of overconfidence in knowledge since, if this tendency is widespread, those researching it could also be affected and may, in view of the nature of the evidence, overstate their case, displaying overconfidence in their findings (Plous, 1993). Consequently, as well as initiatives to enhance the reliability of the study, limitations, and their potential implications for the research findings, are also discussed. Educational research has been criticised because it is too inaccessible and does not assist professional practice (Pring, 2000) and therefore the chapter also addresses the implications of the findings from the study for professional practice, by considering their relevance for professional practitioners responsible for enhancing student learning. As well as discussing what was discovered in the research, the discussion also considers what was not, by clarifying the boundaries of the study and suggesting issues which may merit investigation in future research.

5.2 Aims

In the context of the broader questions regarding self-assessment ability and its implications, the purpose of the research was to investigate self-monitoring, an aspect of metacognitive regulation concerned with the ability to assess one's own learning. It has been argued that improving metacognitive skills can enhance student learning (Veenman, Van Hout-Wolters and Afflerbach, 2006) and consequently, the specific aims of the study were to assess the ability of business students in higher education to monitor their own knowledge accurately, investigate individual differences in this capacity and determine the extent to which it is associated with academic performance. It therefore contributes to knowledge in this field by investigating theory related to the general tendency for overconfidence in knowledge in a specific context: a large UK institution of business higher education.

5.3 Method

The study is also a response to calls for more research in natural settings by focusing on knowledge related to the participants' business study programmes and collecting data in classroom based tests. A number of initiatives were implemented to address suggestions that the reliability of findings in previous studies may have been compromised by respondents' inability to understand the scale used to communicate confidence judgements. These included providing written instructions supported by interactive discussions at data collection sessions, as well eliminating from the sample those who responded inappropriately to a diagnostic device designed to identify misuse of the scale. Additionally, while being informed that the purpose of the research was to assist in understanding influences on student learning, participants were not made aware of its specific focus, to prevent this information biasing their responses. Another problem, in previous studies investigating the association between accurate self-monitoring and academic performance, has been the tendency to operationalise the latter on the basis of knowledge displayed in the test being simultaneously used to determine monitoring accuracy. This approach has been criticised on the grounds that it systematically biases findings in the direction of poor self-monitoring for the less knowledgeable. This study addresses this problem by using independently derived indicators of competence, based

on participants' performance on their study programme. It also explores the association between metaknowledge and academic achievement in greater depth than previous studies, by separately addressing performance in various modes of summative assessment.

5.4 How the Findings Contribute to the Central Theme

In terms of self-monitoring accuracy, the study detected a general tendency for overconfidence in knowledge, which supports the view that overconfidence is a powerful human tendency (Gilfoyle, 2000) and reflects the most commonly reported finding in studies investigating metacognitive self-monitoring, that people tend to overestimate the accuracy of their knowledge (Renner and Renner, 2001). The findings also revealed individual differences, indicating that, as reported in previous studies, overconfidence was greater for males (Acker and Duck, 2008; Pallier, 2003; Barber and Odean, 2001; Lundeberg, Fox, and Punccohar, 1994) and particularly for Chinese students (Culpepper, Zhao and Lowery, 2002; Yates, Lee and Shinotsuka, 1996). They also indicated that overconfidence was greater for older students, a finding contrary to those in previous studies reporting that overconfidence reduces with age (Grimes, 2002; Fischhoff, 1992)

In terms of the potential implications of these findings, Tobias and Everson (2002) suggest that students' ability to distinguish between what they know and what they do not, is an important influence on academic success in all settings. This claim was investigated by exploring the association between overconfidence bias and participants' academic performance on their study programmes. This indicated only a weak, negative association in respect of overall academic performance, a finding reinforced by a lack of evidence that self-monitoring accuracy differed significantly between student groups of differing academic performance. Investigation of differences according to type of assessment task produced similar findings, with weak association detected between overconfidence and each of the two main types of summative assessment used in the institution, examinations and coursework. Similarly, investigation of individual differences in the association between overconfidence and academic performance indicated that while stronger for some groups (e.g. those in their first year at the

university), associations were typically weak. Thus, when considering students' academic performance and using separate instruments to determine overconfidence and academic competence, these findings do not confirm those reported by Koku and Qureshi (2004) in a previous study of business students' examination performance, which detected significantly greater overconfidence for poorer performers.

5.5 Implications for Professional Practice

As discussed in Chapter 1, educational practice tends to be poorly supported by academic research (Hemsley-Brown and Sharp, 2004) and a key factor influencing the adoption of new initiatives is the potential benefits for students (Sparks, 1988). Consequently, this suggests that in terms of student learning, research findings which indicate that an intervention may be associated with higher levels of academic achievement are more likely to impact on professional practice. In respect of metaknowledge, Kennedy, Lawton and Plumlee (2002) argue that professional educators have a responsibility to assist learners in knowing how much they do not know. The positive association between self-monitoring accuracy and academic achievement detected in this study suggests that initiatives to enhance metaknowledge may also improve academic performance. Since business schools aim to equip learners for a career in business it is also important that students appreciate the importance of metathinking skills in improving their chances of success in business (Crittenden and Woodside, 2007). For example, Larres, Ballantine and Whittington (2003) argue that, through stimulating reflection, self-assessment contributes to life-long learning and while they specifically highlight its importance for accountants, who must continually assess their competence during their career, this argument can be extended to include others pursuing a career in business. Therefore, as well as its impact on academic performance, well developed metaknowledge is also important for business education, since it has other potential benefits for business professionals. However, metathinking skills are typically poorly addressed in business schools and while its inclusion in the business curriculum may be advisable, thinking about thinking is a difficult topic to study and students, from undergraduate level up to executives, find studying it challenging (Crittenden and Woodside, 2007). Consequently, Pintrich (2002) argues that, rather than addressing it separately, it should be incorporated in content based

learning activities. However, this is likely to require staff development initiatives since many professional educators have insufficient knowledge of metacognition and, therefore, need the tools to assist them in integrating it into learning activities (Veenman, Van Hout-Wolters and Afflerbach, 2006). However, in the case of metaknowledge however, since their appreciation of the concept may be limited, then prior to providing them with tools to assist in its development in students', it will firstly be necessary to take steps designed to develop their own understanding of the concept.

5.5.1 Engaging Professional Practitioners

Raising higher education professionals' awareness of metaknowledge and the potential implications of its poor development can be accomplished through the dissemination of the findings of this study, accompanied by a review of literature addressing other empirical evidence on metaknowledge. This will occur at the academic institution in which the research was conducted, and at which the researcher is employed, through staff seminars and newsletters and to a wider audience through presentation of the research at academic conferences and publication in academic journals which target professionals working in higher education. The researcher's role as a professional practitioner in higher education should assist in the dissemination of the research in a manner which such practitioners are able to interpret, as recommended by Aram and Salipante (2003).

This will facilitate interventions designed to raise students' awareness of the importance of being able to accurately self-assess the extent of one's own knowledge. However, while awareness alone may help learners to manage overconfidence bias (Russo and Schoemaker, 1992), this alone may be insufficient to do so. Consequently, encouraging education professionals to personalise the issue for students, by providing them with individual feedback in respect of the extent of their own bias, may be beneficial (Fischhoff, 1982) and since it was conducted in a natural setting, the approach adopted in this study is one which can be adapted to do so relatively easily. This therefore addresses Gersten et al's (1997) concern that initiatives implemented in studies conducted in unnatural settings, tend not be adopted in practice unless they can be adapted to reflect naturally occurring environments.

Consequently, having firstly raised the awareness of professional educators working in higher education of the concept of metaknowledge and the potential consequences of its poor development, dissemination of this research will also provide specific guidance on how the tool used in the study can be employed to assess metaknowledge. This will include the manner in which individual questions may be designed as well as instructions which may be provided to students when administering it. Clarification will also be provided on how the bias score may be determined and interpreted. The value of such guidance was highlighted by Gersten et al (1997) when discussing problems encountered in the implementation of self-assessment research into comprehension of written text, which indicated the value of asking students to make predictions and subsequently assess their accuracy. They explain that the impact of this intervention in practice was compromised by poor implementation, with many teachers asking students to provide predictions, but not following this up by subsequently investigating their accuracy. They suggest that this arose due to teachers not fully understanding the importance of different components of this intervention.

Gersten et al (1997) argue that 'top down' approaches to the implementation of new initiatives often fail and highlight the value of dissemination of good practice by educational practitioners communicating details of successful strategies to peers working in the same setting. The members of staff that supported this study are a good target group in this case as Huberman (1990) advocates greater involvement of practitioners in conducting research as a means of enhancing its impact on professional practice. In this research numerous academic staff were involved in the research by initially granting permission to collect data from students studying modules for which they are responsible and subsequently assisting in the compilation of the research instrument used to do so. Involving practitioners in the design of the research in this way can assist in enhancing its practical impact on professional practice (Hemsley-Brown and Sharp, 2004) and therefore implementation of self-assessment procedures by this group initially may be an effective means of later achieving wider dissemination through their communication of the approach to peers. Indeed shortly after completion of the data collection procedure for the study, one member of this group requested support in integrating the activity into the learning programme for the module in question on a more permanent basis for the benefit of future students. Since instruments

have already been designed for the other modules included in this research, the cost of integrating this approach into those modules will also be relatively small and the author intends to pursue this with the staff in question.

5.5.2 Engaging Students

Sternberg (1998) expresses concern that students, who may in the past have been rewarded for more passive learning, may resist the more thoughtful approach that metacognition entails. However, personalising the issue by providing students with individual diagnostic feedback may assist in overcoming this. Multiple-choice tests, based on knowledge related to students' study programmes, could be used to provide formative feedback on the extent of both their knowledge and metaknowledge. The frequent use of these during a study programme would permit repeated feedback which, as discussed earlier, may increase self-assessment accuracy. However, the capacity to administer a series of these tests would be constrained by the time available to design suitable questions and it may be more feasible to introduce them on a more modest scale and develop more gradually over a period of time.

Learners can be more easily engaged in such initiatives by using an approach which is not cognitively over-demanding and adopts a relatively simple scoring system (Farrell and Leung, 2004) and in this case, the bias score used to indicate metaknowledge is not a particularly complex indicator. In terms of the demands of this approach on learners, Swartz (2006) investigated student preferences and found no significant difference between confidence-response multiple-choice questions and the traditional multiple-choice approach, in terms of how difficult they found them to use. He therefore argues that, whether or not practitioners adopt them depends on the trade off between their pedagogic advantages, through better quality, richer feedback and the cost associated with modifying procedures. However, their implementation need not be excessively resource intensive when compared with the use of traditional multiple-choice tests, with the main cost being the time required to generate the bias score from data collected. However this could be facilitated by either training students to do this themselves, or using a computer programme to automate this task. Farrell and Leung (2004) adopted the automated approach, devising a system operated via the internet which could be

used flexibly by students, at a time of their choosing, and provided instant and private feedback. Consequently, using this approach would give practitioners the opportunity to integrate initiatives designed to test for metaknowledge into directed learning activities, rather than conducting them as classroom-based tests.

Personalised diagnostic information provided to business students can be accompanied by information designed to simultaneously develop their appreciation of relevant theory and awareness of the findings from this and other research studies. This can be used to develop their understanding of the potential implications of poor metaknowledge, not only for their future learning, but also for personal and professional decision making. Consequently, the integration of such a learning initiative into business study programmes addresses concerns expressed regarding the inaccessibility of educational research (Hemsley-Brown and Sharp, 2004; Pring, 2000), by engaging both professional practitioners and the learners with whom they engage.

Another initiative which may be effective in promoting the potential value of good metaknowledge to students would be to recognise it through reward. Pintrich (2002) suggests setting learning goals related to metacognitive knowledge and assessing the extent to which these are achieved. In respect of metaknowledge, this could entail rewarding self-monitoring accuracy, as well as more traditionally assessed learning outcomes such as knowledge, application and critical analysis. For example, marks available for a multiple choice examination could be weighted such that students were rewarded for metaknowledge, as indicated by the bias score, as well as knowledge.

5.5.3 Implications for the Author's Own Professional Practice

The author is an academic member of staff at the institution in which the research was conducted, with responsibility for designing, managing and delivering learning activities at undergraduate and post-graduate level to full-time and part-time business students. The theory addressed here, as well as empirical findings derived in this study and previous research, will be incorporated in learning programmes as a means of developing students' understanding of the importance of metaknowledge and its potential implications for learning. This will be achieved by making them aware of the general tendency for overconfidence in knowledge, along with the findings here

indicating its association with academic performance. To further engage learners in addressing this issue, personal diagnostic feedback will also be employed, through the use of the type of multiple choice instruments used in this study to assess both knowledge of issues addressed in their study programme and metaknowledge, as discussed above.

The author also intends to implement initiatives designed to improve knowledge monitoring accuracy. While making individuals aware of their own limitations in this respect can facilitate this, in some cases this may be insufficient (Russo and Schoemaker, 1992) and consequently, other additional interventions will also be implemented. These include explicitly prompting students to consider why their judgements may be incorrect (Koku and Qureshi, 2004; Arkes et al, 1987; Koriat, Lichtenstein and Fischhoff, 1980) and encouraging them to take account of evidence which may challenge their initial beliefs (Hammond, Keeney and Raiffa, 1998; Zakay and Glicksohn, 1992). These will be conducted either as part of directed learning activities using the e-learning platform employed by the university to facilitate student learning or as a classroom-based activity. In the case of the latter, an experimental approach will be adopted in which one sub-group adopts one of these potential remedial strategies while a second control group does not and differences between the two groups in terms of self-assessment accuracy are discussed in the context of literature suggesting the potential value of such remedial initiatives.

At the institution at which he is employed, the author is based in a corporate and management development centre, which designs and delivers interventions to meet the needs of specific business clients. Consequently much of his work is spent designing and delivering learning programmes to facilitate the personal development of aspiring and practicing managers working in a variety of business organisations. This provides the opportunity to address metaknowledge with such individuals and encourage them to reflect on its potential implications for their own practice. The author's own specific contribution to such learning programmes is often in the context of judgement and decision making and therefore, as well as exploring the potential consequences of poor self-monitoring for learning generally, as discussed above, its implications for judgements informing business decisions will also be addressed. The aim of this will be to encourage learners to consider not only implications for their own behaviour, but also

for that of others they may deal with in a professional role as a business manager. For example, as well as appreciating that their decisions may be misinformed, due to over-estimating their own knowledge, it is important that managers are also aware that the same problem may afflict colleagues, on whose judgements they also rely. As discussed earlier, while such individuals should appreciate the importance of metathinking, few business academic programmes incorporate it and business schools have typically not done a good job in integrating it into classroom activities in a manner which will enhance students' appreciation of how metathinking may enhance success in business (Crittenden and Woodside, 2007). The author intends to address this in the context of his own work by building on initiatives he has recently introduced into learning programmes which address behavioural influences on judgement and decision making, by addressing the potential implications of poor metaknowledge for judgement and business decisions.

As well as being able to integrate the specific subject matter addressed in this study into learning activities as discussed above, the author also intends to use his own experience in conducting it to provide more effective guidance in respect of research he himself supervises in his professional role. Participating in both the taught element and thesis stage of the DBA programme has resulted in developing a better personal appreciation of the research process and enhanced the author's understanding of different approaches to research and the merits and potential problems associated with these. Specifically it has also highlighted, or in some cases reinforced, issues such as the importance of being appropriately focused in terms of the aims of research and of seeking feedback from others, as well as the need to ensure that research is theoretically underpinned by a sufficiently critical review of relevant literature. The need to be clear and critically reflective about how methodology and research design impact on research findings is another issue which has been reinforced by the experience of conducting this study. The greater appreciation of the research process which has resulted will enable the author to provide more informed guidance in his role as supervisor to students undertaking their own research, in the form of undergraduate and postgraduate dissertations and research projects as well as doctoral supervision, in which the author will participate in the future.

5.6 Limitations of the Study

This study was informed by realist ontology, empiricist epistemology and quantitative methodology. However, while adopting the view that an objective reality exists that may be empirically investigated, it was acknowledged earlier that it may not have been possible to discover it with certainty, due to limitations associated with the manner in which the research was conducted. These include the difficulty of operationalising a social concept such as overconfidence as well as potential biases on the part of participants as well as the researcher. Research findings may be affected by the role researchers play in generating them and the process in which they reflect on their own actions and values during the research process and their potential impact on the study is known as reflexivity (Robson, 2002). It has been argued that in some cases this has not necessarily been evident and that researchers have been guilty of being uncritical in terms of their position in research (Johnson and Duberley, 2000). Reflexivity may be perceived as being more important in qualitative than quantitative research, since the former values the reflexive 'self awareness of the researcher' (Sarantakos, 2005 p.45). However, while the use of fixed designs in studies such as this usually suggests that researchers act in a detached capacity, their position may nevertheless be influential. Biases may exist not only in methods adopted by researchers, but also as a consequence of their presence in the situation under investigation (Bryman, 2008). For instance, the researcher in this study is a member of academic staff in the institution in which the research took place and therefore, his presence at data collection events may have impacted on the research. Consequently, in order that readers may take them into account when interpreting the findings discussed above, it is important that these, and other limitations, are addressed.

To collect the data necessary to operationalise overconfidence, using the bias score, respondents were required to indicate their confidence in answers to knowledge-based questions. They were given both written and verbal guidance on how to do so and initiatives were implemented to detect inappropriate use of the confidence scale, with respondents excluded from the study accordingly. However, it is possible that, while all included in the study complied with these requirements, some had a clearer understanding of the requirements than others and may have consequently responded differently to equivalent states of uncertainty. An issue worthy of note in this respect is

that English was not the first language of a large proportion of respondents. While the institution's admissions requirements include minimum proficiency levels, to help ensure that students are appropriately equipped for their study programme, language skills in these circumstances will inevitably differ. This was not necessarily problematical in terms of understanding the wording of each individual question, since there was an in-built corrective mechanism to deal with this, in that respondents could have adjusted confidence levels accordingly. For example, faced with a question which they felt they did not clearly understand, they could have provided a correspondingly low confidence level in their chosen answer. However it may have been a more significant issue in terms of understanding the task requirements. To some extent this may have been addressed by the initiatives included in the research design to exclude respondents on the grounds of their inappropriate use of the confidence scale, since these resulted in the exclusion of a greater proportion of non-UK (15%) than UK students (5%). However, it is possible that some respondents with limited understanding of the instructions provided for expressing confidence levels may have evaded detection in this manner.

To enhance the ecological validity of the study, respondents were tested on knowledge related to their study programme. Since they were drawn from across different levels of study, eight variations of the research instrument were used, each focusing on knowledge relevant to that particular group of students. Each was distilled to produce eight sub-sets of questions of approximately equal difficulty, to control for a potential task difficulty confounding effect. However, there may have been other potentially confounding issues arising from the use of different variations of the instrument, which were not addressed. For example, one may have contained more difficult questions earlier in the test than others and if so, this may have been significant. While it would have been possible to control for this using a single test for all participants, it would have not been possible to do so using study programme-based knowledge and consequently this would have compromised the ecological validity of the study. Thus, any limitations associated with the use of different variations of the research instrument may be regarded as a price paid for the ecological validity which many other studies in the field have lacked.

Another potential limitation arises from the extent to which students were motivated to engage sufficiently with the task used for data collection. Some studies, such as Koku and Qureshi's (2004) investigation of business students' performance in a multiple choice examination, have focused on a summative assessment procedure to study confidence in knowledge, whereas in this case, a formative assessment activity was used. Respondents in these circumstances may be less motivated to focus fully on the task in hand and may therefore provide ill-considered responses. The impact of this may have been curtailed to some extent in that those behaving in this way, may have also been more likely to indicate that they did not consent to their data being used in the study and therefore not influenced the findings. However, this may not be the case for all such respondents. While basing the study on a summative assessment activity may have helped to limit this effect, it would have limited it to the investigation of first year undergraduate students since, in the research setting; only these are assessed using multiple-choice based examinations. Another factor which may have assisted in reducing the potential for poorly considered responses by participants is the fact that prior to completing the test, they were informed that they would be provided with information to assist in the interpretation of their results, to allow them to reflect on the potential implications for their future learning. This initiative was designed to encourage them to engage with the task, since providing information in this way can motivate participants to provide considered responses (Davis, 2005).

It has also been suggested that in some circumstances, overconfidence may arise due to the social utility associated with expressions of high confidence levels. In this study, the fact that their responses were provided in an environment which explicitly rewards knowledge may have motivated students to exaggerate their confidence ratings, in order to suggest an air of competence. Submission of their responses to an academic member of staff may have contributed to this, since such staff are responsible for making judgements about their academic performance. While they were assured anonymity and that data would be aggregated for analysis, the request to supply their student registration number may have caused some to mistrust the process and thus contributed to this motivational effect. The more reticent on these grounds may have been excluded from the study due to them not giving consent for their data to be used in the study. However, it is acknowledged that the presence and identity of the researcher may have

influenced participants to some extent and therefore impacted on the data collected and resulting findings. Consequently, if the study was repeated, a more appropriate approach may be to train a research assistant, who is not a member of academic staff at the institution in question, and ask them to brief participants and conduct the data collection procedure.

5.7 Boundaries of the Study and Directions for Future Research

It is important to re-emphasise here that this study did not seek, and makes no claims of evidence for, causal links between overconfidence in knowledge and academic performance. Positive associations between knowledge monitoring accuracy and academic performance do not necessarily indicate that accurate monitoring will result in improved academic achievement. As explained earlier, experimental approaches are more appropriate when aiming to establish causality, but these can be impractical in social research, since manipulating variables, or assigning participants to control groups, may be either unfeasible or unethical (Gill and Johnson, 1991). This study investigated metaknowledge in a natural setting, using a non-experimental approach, which did not entail manipulating variables or using control groups. Its aim was to determine whether associations between overconfidence in knowledge and academic performance exist, which may be indicative of a *possible* causal relationship. The associations reported therefore represent contingent, rather than causal, correlations and while these may not in themselves establish causality, they can play an important role in the identification of causes of events in social science (Kemp and Holmwood, 2003). In this case, associations between overconfidence and academic performance investigated in the study were weak and evident for some sub-groups but not others. Consequently, while by its nature it was unable to detect a causal effect, the study's contribution is the generation of findings which support the view that while students' ability to distinguish between what they know and what they do not may be an influence on academic success in for some, this is not necessarily the case in all settings.

Another important issue to consider is the extent to which the findings from this study can be generalised. Due to manner in which the research was designed, random sampling was not feasible, which compromises its generalisability to some extent.

However, because the sample was representative in respect of the main individual differences investigated, the findings can be reasonably generalised to the population in question (i.e. students studying business at the university in question). While they may also usefully inform those responsible for managing learning in other higher education learning environments, they are not necessarily generalisable to those settings. For example, the study focused on business students and since other subject areas may attract different personality types, who may display differences in behaviour (Koku and Qureshi, 2004), these findings may not be applicable to those studying in other disciplines. Consequently, they may be more generalisable to other business students, studying at different institutions of higher education. However, while learning environments in other institutions may be similar in some respects to that in which this study was conducted, they will inevitably differ to some extent, thus limiting the generalisability of these findings accordingly. Consequently, as Kennedy, Lawton and Plumlee (2002) suggest, understanding of this issue can be enhanced by further research in a variety of institutions and disciplines.

Future studies focusing on business students could address differences by subject area specialism by investigating, for instance, whether findings for those specialising in human resource management differed from those on specialist marketing programmes. This study did not address such differences and it may be, for example, that some groups are more prone to overconfidence in knowledge than others, or that overconfidence is more closely associated with academic performance in some specialist areas of business studies than others. Another question, which could be addressed to develop better understanding of self-monitoring accuracy in business education, is whether the knowledge domain used to determine it is influential. In this study, it was determined in the context of knowledge related to financial and quantitative aspects of study programmes and future research could investigate the extent to which findings may differ if instead, it was assessed based on their marketing knowledge, for example.

Another issue worthy of investigation is the extent to which knowledge monitoring impacts on learning strategies adopted by students and how this relates to academic performance. Hacker et al (2000) suggest that accurate monitoring enables students to use their time more effectively, by focusing on issues in which their knowledge is

lacking and helping them to judge when to finish studying a particular topic. Thus, it can potentially contribute to effective learning and improve academic performance, through better informing the planning aspect of metacognitive regulatory activity. However, while poor appreciation of their own knowledge may constrain students' learning (Pintrich, 2002), metaknowledge may be insufficient in itself to enhance it, since this will also depend on the extent to which it is actually used to inform planning. Investigating the extent to which students' planning decisions are influenced by how well they believe they already know that topic, may therefore help to explain the relatively weak association between self-monitoring accuracy and academic performance detected in this study. This could be achieved by questioning learners about the relative importance of metaknowledge and various other factors which may be influential in planning learning and this is an issue which the author proposes to investigate in future research, in order to further contribute towards better understanding of the implications of individuals' capacity for self-assessment.

Appendices

Appendix 1

Improving Metacognitive Knowledge

A Strategy Evaluation Matrix

Declarative Knowledge	Procedural Knowledge	Conditional Knowledge	
Strategy	***How to Use***	***When to Use***	***Why Use***
Skim	Search for headings, highlighted words, previews, summaries	Prior to reading an extended text.	Provides conceptual overview, helps to focus one's attention.
Slow down	Stop, read, and think.	When information seems especially important.	Enhances focus of one's attention.
Activate prior knowledge	Pause and think about what you already know. Ask what you don't know.	Prior to reading or an unfamiliar task.	Makes new information easier to learn and remember.
Mental integration	Relate main ideas. Use these to construct a theme or conclusion.	When learning complex information or a deeper understanding is needed.	Reduces memory load. Promotes deeper level of understanding.
Diagrams	Identify main ideas, connect them, list supporting details under main ideas, connect supporting details.	When there is a lot of interrelated factual info.	Helps identify main ideas and organize them into categories. Reduces memory load.

Source: Adapted from Schraw (1998 p.120)

Appendix 2

Improving Regulation of Cognition

A Regulatory Checklist

Planning

1. What is the nature of the task?

2. What is my goal?

3. What kind of information and strategies do I need?

4. How much time and resources will I need?

Monitoring

1. Do I have a clear understanding of what I am doing?

2. Does the task make sense?

3. Am I reaching my goals?

4. Do I need to make changes?

Evaluating

1. Have I reached my goal?

2. What worked?

3. What didn't work?

4. Would I do things differently next time?

Source: Schraw (1998 p.121)

Appendix 3

Bloom's Taxonomy of Learning Objectives

Cognitive Domain

Knowledge

1.0 Knowledge

1.10 Knowledge of specifics
- 1.11 Knowledge of terminology
- 1.12 Knowledge of specific facts

1.20 Knowledge of ways and means of dealing with specifics
- 1.21 Knowledge of conventions
- 1.22 Knowledge of trends and sequences
- 1.23 Knowledge of classifications and categories
- 1.24 Knowledge of criteria
- 1.25 Knowledge of methodology

1.30 Knowledge of universals and abstractions in a field
- 1.31 Knowledge of principles and generalizations
- 1.32 Knowledge of theories and structures

Intellectual Abilities and Skills

2.0 Comprehension

- 2.1 Translation
- 2.2 Interpretation
- 2.3 Extrapolation

3.0 Application

4.0 Analysis

- 4.1 Analysis of elements
- 4.2 Analysis of relationships
- 4.3 Analysis of organizational principles

5.0 Synthesis

- 5.1 Production of a unique communication
- 5.2 Production of a plan, or proposed set of operations
- 5.3 Derivation of a set of abstract relations

6.0 Evaluation

- 6.1 Evaluation in terms of internal evidence
- 6.2 Judgements in terms of external criteria

Source: Bloom (1956)

Appendix 4

Revised Taxonomy of Learning Objectives

The Knowledge Dimension	The Cognitive Process Dimension					
	1 Remember	2 Understand	3 Apply	4 Analyse	5 Evaluate	6 Create
A. Factual knowledge						
B. Conceptual knowledge						
C. Procedural knowledge						
D. Meta-cognitive knowledge						

Source: Anderson et al (2001 p.28)

Appendix 5

Revised Taxonomy of Learning Objectives

The Six Categories of the Cognitive Process Dimension

1 Remember

1.1 Recognizing

1.2 Recalling

2 Understand

2.1 Interpreting

2.2 Exemplifying

2.3 Classifying

2.4 Summarizing

2.5 Inferring

2.6 Comparing

2.7 Explaining

3 Apply

3.1 Executing

3.2 Implementing

4 Analyze

4.1 Differentiating

4.2 Organizing

4.3 Attributing

5 Evaluate

5.1 Checking

5.2 Critiquing

6 Create

6.1 Generating

6.2 Planning

6.3 Producing

Source: Anderson et al (2001 p.31)

Appendix 6

Revised Taxonomy of Learning Objectives

The Knowledge Dimension

A. Factual Knowledge

(The basic elements that students must know to be acquainted with a discipline or solve problems in it.)

A_A. Knowledge of terminology

A_B. Knowledge of specific details and elements

B. Conceptual Knowledge

(The interrelationships among the basic elements within a larger structure that enable them to function together.)

B_A. Knowledge of classifications and categories

B_B. Knowledge of principles and generalizations

$B_{C.}$ Knowledge of theories, models, and structures

C. Procedural Knowledge

(How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods.)

C_A. Knowledge of subject-specific skills and algorithms

C_B. Knowledge of subject-specific techniques and methods

C_C. Knowledge of criteria for determining when to use appropriate procedures

D. Metacognitive Knowledge

(Knowledge of cognition in general as well as awareness and knowledge of one's own cognition.)

D_A. Strategic knowledge

D_B. Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge

D_C. Self-knowledge

Source: Anderson et al (2001 p.29)

Appendix 7

Test for Overconfidence Using Continuous Variables (Range Questions)

For each of the following questions, make a low and a high estimate for which you believe there is a 90% chance that the correct answer lies between your two estimated figures.

Your challenge is to try to ensure that the difference between your two estimates for each question is neither too high nor too low.

		90% Confidence Range	
		Low Number	***High Number***
1	What is the length of the River Amazon in kilometres?		
2	How many weeks did Beatles recordings spend at number one in the 1960s?		
3	How many bones are there in the human body?		
4	How old was Albert Einstein when he died?		
5	What is the total population of Australia?		
6	What is the height of Mount Everest in metres?		
7	What is the east to west width of the Pacific Ocean in kilometres?		
8	What is the average life span of a giraffe in months?		
9	What is the distance between Edinburgh and London in kilometres?		
10	How many weeks did Bill Clinton spend as president of the USA?		

Source: Adapted from Russo and Schoemaker (1989)

Appendix 8

The Research Instrument

TESTING YOUR UNDERSTANDING

Instructions

- You are required to answer multiple-choice questions in which you indicate your response by ticking one of the four alternatives provided.

Confidence Levels

- After each question you are also required to record your confidence level, indicating the % chance that you believe you have answered correctly.

- Your confidence level for each question should fall between 25% and 100%.

If you are certain that your answer is correct,

You should record a 100% response.

If your answer is a complete guess

As there are four possible answers to choose from, you have a 1 in 4 chance of being correct and you should therefore record 25%.

In other cases

You should provide a response somewhere between 25% and 100% to indicate your confidence in the choice you have made.
(If for example you believe that you are certain that two of the possible answers are definitely incorrect but you believe that the other two have an equal chance of being correct, you should choose one and indicate a 50% confidence rating)

Example

Which of the following countries is biggest in terms of area?

a)	*Peru*	☐
b)	*Mexico*	☐
c)	*Denmark*	☐
d)	*Italy*	✓

Confidence level (25-100%) …60……..%

This would indicate that you believe there is a 60% chance that your judgement that Italy is the biggest country is correct (ie. if you were answering 10 questions with this level of confidence you would expect to be correct 6 times)

- You should work through the questions as quickly and accurately as you can.

- Data collected in this exercise will be used to assist you in appreciating issues addressed to date in the module and you will be provided with individual feedback on your responses. It will also be used for research purposes to assist in understanding influences on student learning. In this respect results generated, which will be analysed in conjunction with student academic performance on the programme, will be anonymous and aggregated for research purposes.
  Y**our name will not be used in any part of the research and your responses in this activity will not affect your mark for the module.**

  Please tick here to indicate whether you agree to this data being used in this way for research purposes.

  I agree ☐

  I disagree ☐

Submission of Questionnaire

Should you be willing for this data to be incorporated into the research findings, please complete the following details by ticking the relevant box.

General Details

Office use

1) Age 18-21 ☐ 22-25 ☐ 26-29 ☐ 30-33 ☐ 34+ ☐ *1*……

2) Gender *2*……

Male ☐

Female ☐

3) Country of origin…… 3……

4) Student number…….. 4……

5) On which course are you currently studying?

…….. *5*……

QUESTIONS

1) The Bretton Woods System was:

a) A floating exchange rate system ☐

b) A fixed exchange rate system ☐

c) A managed floating exchange rate system ☐

d) A sinking exchange rate system ☐

Confidence level (25-100%) …….%

---

2) According to the Anglo American model the objective of a company is to:

a) Maximise corporate wealth ☐

b) Maximise sales income ☐

c) Maximise profit ☐

d) Maximise shareholder wealth ☐

Confidence level (25-100%) …….%

3) Fixed exchange rates promote international trade by:

a) Increasing exchange rate risk ☐

b) Increasing the possibility of currency speculation ☐

c) Increasing the productivity of companies ☐

d) Removing exchange rate risk ☐

Confidence level (25-100%) …….%

---

4) The Asian currency crisis occurred in:

a) 1997 ☐

b) 1967 ☐

c) 1977 ☐

d) 1987 ☐

Confidence level (25-100%) …….%

5) NAFTA stands for

a) North Atlantic Full Trading Area ☐

b) New Approach to Funding Trading Associations ☐

c) North American Free Trade Agreement ☐

d) New Association for Funding Trade Alliances ☐

Confidence level (25-100%) …….%

---

6) Exchange rate risk for a Multi-National Company arises because:

a) Overseas cash flows may depreciate against the home currency ☐

b) Costs in the home country may rise ☐

c) Income earned in the home country may fall ☐

d) Overseas cash flows may appreciate against the home currency ☐

Confidence level (25-100%) …….%

7) The Bretton Woods System collapsed in:

a) 1951 ☐

b) 1961 ☐

c) 1971 ☐

d) 1981 ☐

Confidence level (25-100%) …….%

---

8) The Maastricht convergence criteria in respect of the European single currency included which of the following
:

a) Inflation rates ☐

b) Growth rates ☐

c) Unemployment rates ☐

d) Wage rates ☐

Confidence level (25-100%) …….%

9) Which of the following represents the greatest economic and political integration for trading blocs?

a) Common market ☐

b) Economic union ☐

c) Free-trade area ☐

d) Customs union ☐

Confidence level (25-100%) …….%

---

10) A rational bubble

a) Exchange rates adjust to their correct level ☐

b) Exchange rates stay at an incorrect level ☐

c) Speculators hold an undervalued currency ☐

d) Speculators will be unaware that a problem is developing ☐

Confidence level (25-100%) …….%

11) After the financial liberalizations in the 1980's:

a) Currency crises have not occurred ☐

b) Currency crises have been accompanied by banking crises ☐

c) Banking crises have not occurred ☐

d) Currency crises have not been accompanied by banking crises ☐

Confidence level (25-100%) …….%

---

12) Ethnocentrism is the belief that:

a) Businesses should seek ways to minimise risk ☐

b) Strategies for change should be initiated by government ☐

c) Financial systems should be closely regulated ☐

d) One's own culture is superior to others ☐

Confidence level (25-100%) …….%

13) A chaebol is:

a) A financial instrument ☐

b) A state owned conglomerate ☐

c) A type of business takeover ☐

d) A family owned conglomerate ☐

Confidence level (25-100%) …….%

---

14) The term ‘contagion’ is used to describe:

a) The financial objectives of companies ☐

b) Ethical concerns over business managers ☐

c) A type of financial panic ☐

d) Growth rates in business income ☐

Confidence level (25-100%) …….%

15) What was the total turnover of British Airways in the year 2004-05?

a) £7,811m. ☐

b) £7,812m. ☐

c) £7,813m. ☐

d) £7,814m. ☐

Confidence level (25-100%) …….%

---

16) Which of the following countries did not adopt the single European Currency in 1999?

a) Ireland ☐

b) Austria ☐

c) Belgium ☐

d) Switzerland ☐

Confidence level (25-100%) …….%

17) Which of the following is **not** a learning outcome for this module?

a) Identify the issues that influence overseas investment decisions ☐

b) Critically discuss ways of raising international funds ☐

c) Identify characteristics of exchange rate systems ☐

d) Critically discuss overseas investment decisions made by multinational businesses ☐

Confidence level (25-100%) …….%

---

18) Protectionism is defined by Eitman, Stonehill and Moffet as:

a) An attempt by a company to reduce its tax liability ☐

b) An attempt by a company to reduce the risks to which it is exposed ☐

c) An attempt by government to reduce corruption ☐

d) An attempt by a government to protect particular industries from foreign competition ☐

Confidence level (25-100%) …….%

19) Interest Rate Parity (IRP) provides the linkage between:

a) Interest rates and inflation rates ☐

b) Foreign exchange markets and inflation rates ☐

c) Inflation rates and international money markets ☐

d) Foreign exchange markets and international money markets ☐

Confidence level (25-100%) …….%

---

20) Which of the following type of risks faced by Multi National Enterprises is classed as firm-specific?

a) Foreign exchange risks ☐

b) Transfer risks ☐

c) Cultural risk ☐

d) International risk ☐

Confidence level (25-100%) …….%

21) University regulations on plagiarism mean that students should

a) Only use their own opinions when writing assignments ☐

b) Never use direct quotes from other people ☐

c) Never discuss assignments with other students ☐

d) Provide references when using other people's ideas or words ☐

Confidence level (25-100%) …….%

---

22) The balance of payments is a summary of:

a) All international transactions between a country and all other countries ☐

b) All international transactions between two companies ☐

c) All costs incurred by a company when trading with foreign-owned businesses ☐

d) All costs incurred by all companies in a country when trading with foreign companies ☐

Confidence level (25-100%) …….%

23) In the late 1990s Daewoo had a debt to equity ratio which:

a) Was significantly less than the average of U.S. companies ☐

b) Was significantly higher than the average of U.S. companies ☐

c) Was significantly lower than that of similar Korean conglomerates ☐

d) Was approximately equal to that of similar Korean conglomerates ☐

Confidence level (25-100%) …….%

---

24) When forecasting using Purchasing Power Parity, the UK forward rate equals:

a) Spot rate x $\frac{1 + \text{overseas inflation rate}}{1 + \text{UK inflation rate}}$ ☐

b) Spot rate x $\frac{1 - \text{overseas inflation rate}}{1 - \text{UK inflation rate}}$ ☐

c) Spot rate x $\frac{1 + \text{UK inflation rate}}{1 + \text{overseas inflation rate}}$ ☐

d) Spot rate x $\frac{1 - \text{UK inflation rate}}{1 - \text{overseas inflation rate}}$ ☐

Confidence level (25-100%) …….%

25) Which of the following countries had the highest score in Transparency International's Corruption Perception index for 2001?

a) France ☐

b) Canada ☐

c) Turkey ☐

d) Mexico ☐

Confidence level (25-100%) …….%

---

26) According to Conklin (2002) which of the following is suggested as a method for dealing with exchange rate risks?

a) Borrow in a different currency to the one in which the business is trading ☐

b) Invest in as many different countries as possible ☐

c) Arrange for most of the investment cost to be incurred over as short a period as possible ☐

d) Borrow domestically to do business domestically ☐

Confidence level (25-100%) …….%

27) ASEAN stands for:

a) Association of Southern European Allied Nations ☐

b) Association of Southeast Asian Nations ☐

c) Association of Southeast African Nations ☐

d) Association of South Eastern Atlantic Nations ☐

Confidence level (25-100%) …….%

---

28) According to the World Almanac and Book of Facts 2002, which of the following languages, has the highest number of native speakers:

a) Bengali ☐

b) Japanese ☐

c) French ☐

d) Spanish ☐

Confidence level (25-100%) …….%

29) Which of the following type of risks faced by Multi National Enterprises is classed as country-specific?

a) Governance risks ☐

b) Business risks ☐

c) Cultural and international risk ☐

d) Foreign exchange risks ☐

Confidence level (25-100%) …….%

---

30) Which of the following is acknowledged by Yoshino as typical of Japanese work customs:

a) High employee turnover ☐

b) Very little socialising after work ☐

c) Decisions taken primarily by senior managers ☐

d) Group based rewards ☐

Confidence level (25-100%) …….%

31) Which of the following countries ranked highest in the United Nations 2002 Human Development Index?

a) Japan ☐

b) France ☐

c) Brazil ☐

d) Germany ☐

Confidence level (25-100%) …….%

Appendix 9

Coding Plan

Variable	Description	Code
Age	18-21	1
	22-25	2
	26-29	3
	30-33	4
	34+	5
Gender	male	1
	female	2
Study Programme	babs	1
	babm	2
	bwhrm	3
	bwm	4
	bwim	5
	bwe	6
	mwbm	7
	baac	8
	bwf	9
	baaf	10
	fim	11
	bacm	12
	mba	13
	magfm	14
	mabm	15
	maiba	16
	mabfm	17
	baibs	18
	baiba	19
	bwit	20
	bwlscm	21
	bamkt	22
	bw tour	23
Country of Origin	uk	1
	china	2
	indonesia	3
	thailand	4
	vietnam	5
	sweden	6
	india	7
	ukraine	8
	nepal	9
	taiwan	10

	cyprus	11
	nigeria	12
	russia	13
	france	14
	germany	15
	liberia	16
	rwanda	17
	spain	18
	iran	19
	tanzania	20
	kuwait	21
	hong kong	22
	malaysia	23
	bangladesh	24
	SouthAfrica	25
	angola	26
	sri lanka	27
	zimbabwe	28
	bahrain	29
	brunei	30
	uae	31
	slovakia	32
	egypt	33
	latvia	34
	ireland	35
	Saudi	36
	gibraltar	37
	congo	38
	estonia	39
	bahrain	40
	lithuania	41
	denmark	42
	gambia	43
Module	BM0118	2
	FN0134	3
	FN0205	4
	MN0253	5
	MN0307	6
	MN0353	7
	MN0431	8
	MN0457	9
Agreement to participate	yes	1
	no	2
Appropriate use of scale	yes	1
	no	2

Appendix 10

Analysis of Differences in Mean Knowledge Scores

30 Item Test Instruments

Knowledge Score by Module Code

Module Code	Mean	N
BM0118	.4739	70
FN0134	.4838	57
FN0205	.4242	22
MN0253	.4663	158
MN0307	.5519	27
MN0353	.5620	47
MN0431	.5096	85
MN0457	.5198	42
Total	.4925	508

Test of Homogeneity of Variances

Knowledge Score

Levene Statistic	df1	df2	Sig.
3.960	7	500	.000

Kruskal-Wallis Test

Ranks

	Module Code	N	Mean Rank
Knowledge Score	BM0118	70	233.42
	FN0134	57	251.74
	FN0205	22	178.39
	MN0253	158	226.79
	MN0307	27	335.65
	MN0353	47	315.01
	MN0431	85	271.50
	MN0457	42	283.20
	Total	508	

Test Statistics[a,b]

	Knowledge Score
Chi-Square	32.128
df	7
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: Module Code

Appendix 11

Analysis of Differences in Mean Knowledge Scores

Randomly Sampled 25 Test Items

Means by Module Code

Module Code	Mean	N
BM0118	.5026	70
FN0134	.5042	57
FN0205	.4818	22
MN0253	.4960	158
MN0307	.5107	27
MN0353	.5245	47
MN0431	.5053	85
MN0457	.4969	42
Total	.5023	508

Kruskal-Wallis Test

Ranks

	Module Code	N	Mean Rank
Knowledge Score	BM0118	70	254.10
	FN0134	57	262.56
	FN0205	22	233.48
	MN0253	158	249.71
	MN0307	27	272.06
	MN0353	47	266.51
	MN0431	85	254.19
	MN0457	42	249.15
	Total	508	

Test Statistics[a,b]

	Knowledge Score
Chi-Square	1.560
Df	7
Asymp. Sig.	.980

a. Kruskal Wallis Test

b. Grouping Variable: Module Code

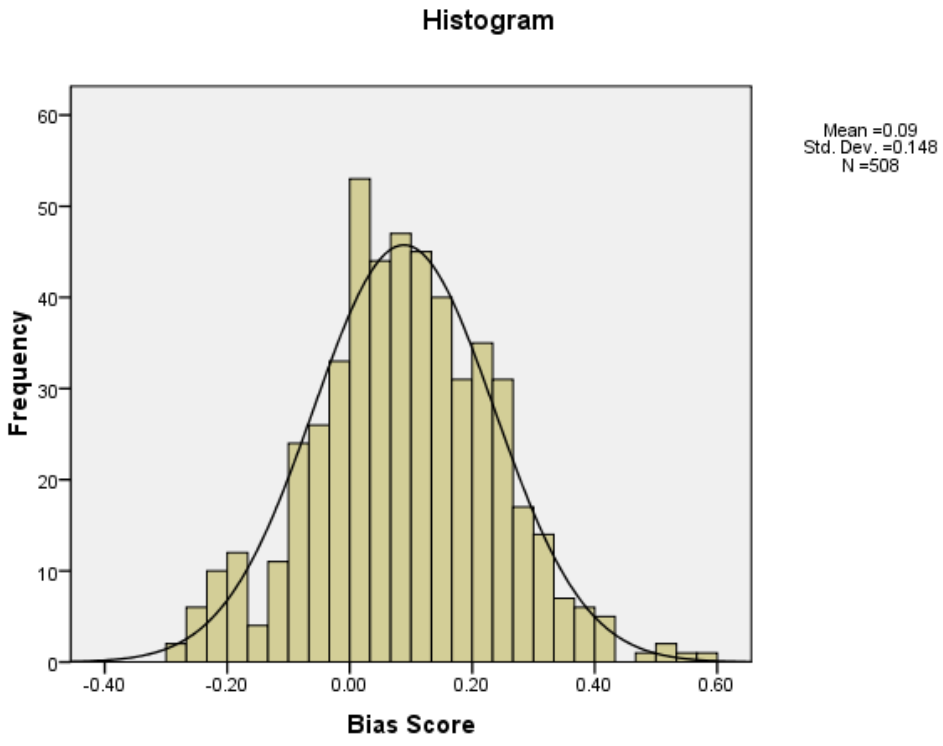
Appendix 12

Test for Normal Distribution of the Bias Score

One-Sample Kolmogorov-Smirnov Test

		Bias Score
N		508
Normal Parameters[a]	Mean	.0885
	Std. Deviation	.14765
Most Extreme Differences	Absolute	.034
	Positive	.034
	Negative	-.027
Kolmogorov-Smirnov Z		.762
Asymp. Sig. (2-tailed)		.606

a. Test distribution is Normal.



Appendix 13

Bias Score – One Sample t-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Bias Score	508	.0885	.14765	.00655

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Bias Score	13.504	507	.000	.08846	.0756	.1013

Sig. (1-tailed) = .000/2= 0.000

Appendix 14

Correlation between Bias Score and Knowledge Score

Correlations

		Bias Score	Knowledge Score
Bias Score	Pearson Correlation	1.000	-.504**
	Sig. (1-tailed)		.000
	N	508	508
Knowledge Score	Pearson Correlation	-.504**	1.000
	Sig. (1-tailed)	.000	
	N	508	508

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 15

Correlation between Mean Confidence and Bias Score

Correlations

		Mean Confidence	Bias Score
Mean Confidence	Pearson Correlation	1.000	.561**
	Sig. (1-tailed)		.000
	N	508	508
Bias Score	Pearson Correlation	.561**	1.000
	Sig. (1-tailed)	.000	
	N	508	508

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 16

Test for Age Differences in Bias Score

Group Statistics

	Age Group	N	Mean	Std. Deviation	Std. Error Mean
Bias Score	18-21 years	297	.0725	.13849	.00804
	22+ years	207	.1125	.15728	.01093

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Bias Score	Equal variances assumed	2.245	.135	-3.020	502	.003	-.04005	.01326	-.06611	-.01399
	Equal variances not assumed			-2.952	406.222	.003	-.04005	.01357	-.06673	-.01338

Appendix 17

Test for Gender Differences in Bias Score

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Bias Score	Male	281	.0986	.14723	.00878
	Female	224	.0761	.14844	.00992

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Bias Score	Equal variances assumed	.000	.992	1.700	503	.090	.02250	.01324	-.00351	.04850
	Equal variances not assumed			1.698	476.509	.090	.02250	.01325	-.00354	.04853

Sig. (1-tailed) = .090/2= 0.045

Appendix 18

Bias Score – Test for Interaction between Age and Gender

Descriptive Statistics

Dependent Variable:Bias Score

Gender	Age Group	Mean	Std. Deviation	N
Male	18-21 years	.0776	.13903	169
	22+ years	.1332	.15169	111
	Total	.0996	.14648	280
Female	18-21 years	.0661	.13851	127
	22+ years	.0885	.16098	96
	Total	.0757	.14867	223
Total	18-21 years	.0726	.13869	296
	22+ years	.1125	.15728	207
	Total	.0890	.14779	503

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Bias Score

F	df1	df2	Sig.
.993	3	499	.396

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Gender + over21 + Gender * over21

Tests of Between-Subjects Effects

Dependent Variable:Bias Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.306[a]	3	.102	4.778	.003
Intercept	4.020	1	4.020	188.220	.000
Gender	.095	1	.095	4.454	.035
Age	.184	1	.184	8.608	.004
Gender * Age	.033	1	.033	1.553	.213
Error	10.658	499	.021		
Total	14.952	503			
Corrected Total	10.964	502			

a. R Squared = .028 (Adjusted R Squared = .022)

Appendix 19

Test for Differences in Bias Score between UK and Chinese Students

Group Statistics

	Country of Origin	N	Mean	Std. Deviation	Std. Error Mean
Bias Score	UK	269	.0509	.13124	.00800
	China	102	.1617	.15159	.01501

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Bias Score	Equal variances assumed	1.487	.223	-6.950	369	.000	-.11081	.01594	-.14216	-.07946
	Equal variances not assumed			-6.515	161.644	.000	-.11081	.01701	-.14440	-.07722

Sig. (1-tailed) = .000/2= 0.000

Appendix 20

Bias Score - Test for Interaction between Country and Age

Descriptive Statistics

Dependent Variable:Bias Score

Country of Origin	Age Group	Mean	Std. Deviation	N
UK	18-21 years	.0520	.12898	235
	22+ years	.0429	.14776	34
	Total	.0509	.13124	269
China	18-21 years	.1996	.15894	25
	22+ years	.1492	.14908	76
	Total	.1617	.15235	101
Total	18-21 years	.0662	.13882	260
	22+ years	.1164	.15600	110
	Total	.0811	.14575	370

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Bias Score

F	df1	df2	Sig.
.694	3	366	.556

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Origin + Age + Origin * Age

Tests of Between-Subjects Effects

Dependent Variable:Bias Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.952[a]	3	.317	16.867	.000
Intercept	2.268	1	2.268	120.533	.000
Origin	.742	1	.742	39.450	.000
Age	.041	1	.041	2.163	.142
Origin * Age	.020	1	.020	1.046	.307
Error	6.887	366	.019		
Total	10.273	370			
Corrected Total	7.839	369			

a. R Squared = .121 (Adjusted R Squared = .114)

Appendix 21

Bias Score - Test for Interaction between Country and Gender

Descriptive Statistics

Dependent Variable:Bias Score

Country of Origin	Gender	Mean	Std. Deviation	N
UK	Male	.0702	.13822	159
	Female	.0229	.11595	109
	Total	.0510	.13147	268
China	Male	.1505	.16144	42
	Female	.1695	.14518	60
	Total	.1617	.15159	102
Total	Male	.0870	.14665	201
	Female	.0750	.14489	169
	Total	.0815	.14577	370

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Bias Score

F	df1	df2	Sig.
1.814	3	366	.144

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Origin + Gender + Origin * Gender

Tests of Between-Subjects Effects

Dependent Variable:Bias Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.059[a]	3	.353	19.042	.000
Intercept	3.051	1	3.051	164.617	.000
Origin	.920	1	.920	49.642	.000
Gender	.014	1	.014	.769	.381
Origin * Gender	.079	1	.079	4.237	.040
Error	6.783	366	.019		
Total	10.298	370			
Corrected Total	7.841	369			

a. R Squared = .135 (Adjusted R Squared = .128)

Appendix 22

Test for Gender Differences in Bias Score – Chinese Students

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Bias Score	Male	42	.1505	.16144	.02491
	Female	60	.1695	.14518	.01874

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Bias Score	Equal variances assumed	.349	.556	-.622	100	.535	-.01902	.03059	-.07972	.04167
	Equal variances not assumed			-.610	82.242	.543	-.01902	.03117	-.08104	.04299

Appendix 23

Test for Differences in Bias Score by Level of Study

Report

Level of Study		Mean Confidence	Knowledge Score	Bias Score
4	Mean	.5968	.5033	.0934
	N	127	127	127
5	Mean	.5529	.4943	.0584
	N	180	180	180
6	Mean	.5819	.5195	.0623
	N	74	74	74
7	Mean	.6439	.5025	.1413
	N	127	127	127
Total	Mean	.5908	.5023	.0885
	N	508	508	508

Test of Homogeneity of Variances

Bias Score

Levene Statistic	df1	df2	Sig.
.667	3	504	.573

ANOVA

Bias Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.571	3	.190	9.152	.000
Within Groups	10.482	504	.021		
Total	11.053	507			

Multiple Comparisons

Bias Score

Tukey HSD

(I) Level of Study	(J) Level of Study	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
4	5	.03494	.01671	.158	-.0081	.0780
	6	.03109	.02109	.454	-.0233	.0855
	7	-.04795*	.01810	.041	-.0946	-.0013
5	4	-.03494	.01671	.158	-.0780	.0081
	6	-.00385	.01991	.997	-.0552	.0475
	7	-.08289*	.01671	.000	-.1260	-.0398
6	4	-.03109	.02109	.454	-.0855	.0233
	5	.00385	.01991	.997	-.0475	.0552
	7	-.07904*	.02109	.001	-.1334	-.0247
7	4	.04795*	.01810	.041	.0013	.0946
	5	.08289*	.01671	.000	.0398	.1260
	6	.07904*	.02109	.001	.0247	.1334

*. The mean difference is significant at the 0.05 level.

Appendix 24

Bias Score - Test for Interaction between Level of Study and Age

Descriptive Statistics

Dependent Variable:Bias Score

Level of Study	Age Group	Mean	Std. Deviation	N
4	18-21 years	.0922	.13652	114
	22+ years	.1038	.19658	13
	Total	.0934	.14285	127
5	18-21 years	.0545	.14271	158
	22+ years	.0833	.12285	21
	Total	.0579	.14052	179
6	18-21 years	.0772	.10742	18
	22+ years	.0620	.14136	55
	Total	.0658	.13325	73
7	18-21 years	.1443	.08696	7
	22+ years	.1422	.15981	118
	Total	.1423	.15641	125
Total	18-21 years	.0725	.13849	297
	22+ years	.1125	.15728	207
	Total	.0889	.14767	504

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Bias Score

F	df1	df2	Sig.
1.363	7	496	.219

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Level + Age + Level * Age

Tests of Between-Subjects Effects

Dependent Variable:Bias Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.591[a]	7	.084	4.034	.000
Intercept	1.582	1	1.582	75.608	.000
Level	.130	3	.043	2.064	.104
Age	.001	1	.001	.070	.791
Level * Age	.016	3	.005	.257	.856
Error	10.378	496	.021		
Total	14.953	504			
Corrected Total	10.969	503			

a. R Squared = .054 (Adjusted R Squared = .041)

Appendix 25

Bias Score - Test for Interaction between Level of Study and Gender

Descriptive Statistics

Dependent Variable:Bias Score

Level of Study	Gender	Mean	Std. Deviation	N
4	Male	.1076	.14285	74
	Female	.0746	.14301	52
	Total	.0940	.14327	126
5	Male	.0547	.13975	98
	Female	.0629	.14176	82
	Total	.0584	.14033	180
6	Male	.0662	.12502	39
	Female	.0580	.14830	35
	Total	.0623	.13563	74
7	Male	.1687	.14802	70
	Female	.1087	.16165	55
	Total	.1423	.15641	125
Total	Male	.0986	.14723	281
	Female	.0761	.14844	224
	Total	.0886	.14805	505

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Bias Score

F	df1	df2	Sig.
.489	7	497	.843

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Level + Gender + Level * Gender

Tests of Between-Subjects Effects

Dependent Variable:Bias Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.727[a]	7	.104	5.005	.000
Intercept	3.469	1	3.469	167.067	.000
Level	.519	3	.173	8.338	.000
Gender	.061	1	.061	2.928	.088
Level * Gender	.092	3	.031	1.483	.218
Error	10.319	497	.021		
Total	15.014	505			
Corrected Total	11.047	504			

a. R Squared = .066 (Adjusted R Squared = .053)

Appendix 26

Bias Score - Test for Interaction between Level of Study and Country

Descriptive Statistics

Dependent Variable:Bias Score

Level of Study	Country of Origin	Mean	Std. Deviation	N
4	UK	.0749	.12909	94
	China	.1308	.16938	13
	Total	.0817	.13492	107
5	UK	.0326	.12810	145
	China	.2167	.12196	21
	Total	.0559	.14103	166
6	UK	.0205	.13442	20
	China	.1163	.13704	19
	Total	.0672	.14242	39
7	UK	.1500	.12508	10
	China	.1639	.15982	49
	Total	.1615	.15360	59
Total	UK	.0509	.13124	269
	China	.1617	.15159	102
	Total	.0813	.14561	371

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Bias Score

F	df1	df2	Sig.
.532	7	363	.810

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Level + Origin + Level * Origin

Tests of Between-Subjects Effects

Dependent Variable:Bias Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.243[a]	7	.178	9.760	.000
Intercept	2.246	1	2.246	123.506	.000
Level	.156	3	.052	2.863	.037
Origin	.335	1	.335	18.406	.000
Level * Origin	.212	3	.071	3.884	.009
Error	6.602	363	.018		
Total	10.298	371			
Corrected Total	7.845	370			

a. R Squared = .158 (Adjusted R Squared = .142)

Appendix 27

Test for Differences in Bias Score by Entry Status

Level 6 Participants

Report

Entry Status		Mean Confidence	Knowledge Score	Bias Score
Direct Entrants	Mean	.6086	.5198	.0886
	N	44	44	44
Continuing Students	Mean	.5427	.5190	.0237
	N	30	30	30
Total	Mean	.5819	.5195	.0623
	N	74	74	74

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Bias Score	Equal variances assumed	.675	.414	2.068	72	.042	.06497	.03142	.00234	.12760
	Equal variances not assumed			2.161	70.284	.034	.06497	.03007	.00500	.12494

Appendix 28

Test for Differences in Bias Score by Entry Status

Level 7 Participants

Report

Entry Status		Mean Confidence	Knowledge Score	Bias Score
Direct Entrants	Mean	.6350	.5132	.1219
	N	80	80	80
Continuing Students	Mean	.6589	.4843	.1745
	N	47	47	47
Total	Mean	.6439	.5025	.1413
	N	127	127	127

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Bias Score	Equal variances assumed	2.324	.130	-1.859	125	.065	-.05259	.02829	-.10858	.00339
	Equal variances not assumed			-1.957	111.744	.053	-.05259	.02688	-.10585	.00066

Appendix 29

Correlation between Bias Score and Academic Performance

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.113**	-.096*	.045	-.138**	-.004
	Sig. (1-tailed)		.007	.019	.348	.001	.479
	N	508	469	463	78	467	152
Overall Mark	Pearson Correlation	-.113**	1.000	.867**	.182	.855**	.739**
	Sig. (1-tailed)	.007		.000	.056	.000	.000
	N	469	469	463	78	467	152
Exam Mark	Pearson Correlation	-.096*	.867**	1.000	.139	.683**	.402**
	Sig. (1-tailed)	.019	.000		.112	.000	.000
	N	463	463	463	78	461	148
Presentation Mark	Pearson Correlation	.045	.182	.139	1.000	.236*	-.105
	Sig. (1-tailed)	.348	.056	.112		.019	.372
	N	78	78	78	78	78	12
Coursework Mark	Pearson Correlation	-.138**	.855**	.683**	.236*	1.000	.681**
	Sig. (1-tailed)	.001	.000	.000	.019		.000
	N	467	467	461	78	467	152
Dissertation Mark	Pearson Correlation	-.004	.739**	.402**	-.105	.681**	1.000
	Sig. (1-tailed)	.479	.000	.000	.372	.000	
	N	152	152	148	12	152	152

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

Appendix 30

Test for Differences in Bias Score by Academic Performance Group

Descriptives

Bias Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Low	156	.1049	.14899	.01193	.0814	.1285	-.28	.42
Middle	156	.0960	.15303	.01225	.0718	.1202	-.25	.54
High	157	.0703	.14162	.01130	.0479	.0926	-.28	.59
Total	469	.0904	.14836	.00685	.0769	.1038	-.28	.59

Test of Homogeneity of Variances

Bias Score

Levene Statistic	df1	df2	Sig.
1.286	2	466	.277

ANOVA

Bias Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.102	2	.051	2.321	.099
Within Groups	10.200	466	.022		
Total	10.301	468			

Appendix 31

Correlation between Bias Score and Academic Performance

Undergraduate Students – Study Level 4

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1	-.188*	-.156*	.092	-.212*	.[a]
	Sig. (1-tailed)		.021	.047	.280	.011	.
	N	127	117	117	42	117	0
Overall Mark	Pearson Correlation	-.188*	1	.946**	.242	.730**	.[a]
	Sig. (1-tailed)	.021		.000	.062	.000	.
	N	117	117	117	42	117	0
Exam Mark	Pearson Correlation	-.156*	.946**	1	.092	.563**	.[a]
	Sig. (1-tailed)	.047	.000		.281	.000	.
	N	117	117	117	42	117	0
Presentation Mark	Pearson Correlation	.092	.242	.092	1	.356*	.[a]
	Sig. (1-tailed)	.280	.062	.281		.010	.
	N	42	42	42	42	42	0
Coursework Mark	Pearson Correlation	-.212*	.730**	.563**	.356*	1	.[a]
	Sig. (1-tailed)	.011	.000	.000	.010		.
	N	117	117	117	42	117	0
Dissertation Mark	Pearson Correlation	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]
	Sig. (1-tailed)	.	.	.	.	.	
	N	0	0	0	0	0	0

*. Correlation is significant at the 0.05 level (1-tailed).

a. Cannot be computed because at least one of the variables is constant.

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 32

Correlation between Bias Score and Academic Performance

Undergraduate Students – Study Level 5

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1	-.055	-.051	-.147	-.111	.[a]
	Sig. (1-tailed)		.247	.262	.334	.085	.
	N	180	156	156	11	154	0
Overall Mark	Pearson Correlation	-.055	1	.860[**]	.601[*]	.862[**]	.[a]
	Sig. (1-tailed)	.247		.000	.025	.000	.
	N	156	156	156	11	154	0
Exam Mark	Pearson Correlation	-.051	.860[**]	1	.610[*]	.652[**]	.[a]
	Sig. (1-tailed)	.262	.000		.023	.000	.
	N	156	156	156	11	154	0
Presentation Mark	Pearson Correlation	-.147	.601[*]	.610[*]	1	.531[*]	.[a]
	Sig. (1-tailed)	.334	.025	.023		.046	.
	N	11	11	11	11	11	0
Coursework Mark	Pearson Correlation	-.111	.862[**]	.652[**]	.531[*]	1	.[a]
	Sig. (1-tailed)	.085	.000	.000	.046		.
	N	154	154	154	11	154	0
Dissertation Mark	Pearson Correlation	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]
	Sig. (1-tailed)	.	.	.	.	.	
	N	0	0	0	0	0	0

a. Cannot be computed because at least one of the variables is constant.

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

Appendix 33

Correlation between Bias Score and Academic Performance

Undergraduate Students – Study Level 6

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.071	-.043	.425	-.129	-.008
	Sig. (1-tailed)		.276	.358	.097	.138	.479
	N	74	73	73	11	73	46
Overall Mark	Pearson Correlation	-.071	1.000	.899**	.062	.919**	.810**
	Sig. (1-tailed)	.276		.000	.428	.000	.000
	N	73	73	73	11	73	46
Exam Mark	Pearson Correlation	-.043	.899**	1.000	.031	.742**	.582**
	Sig. (1-tailed)	.358	.000		.464	.000	.000
	N	73	73	73	11	73	46
Presentation Mark	Pearson Correlation	.425	.062	.031	1.000	-.034	.[a]
	Sig. (1-tailed)	.097	.428	.464		.461	.
	N	11	11	11	11	11	1
Coursework Mark	Pearson Correlation	-.129	.919**	.742**	-.034	1.000	.826**
	Sig. (1-tailed)	.138	.000	.000	.461		.000
	N	73	73	73	11	73	46
Dissertation Mark	Pearson Correlation	-.008	.810**	.582**	.[a]	.826**	1.000
	Sig. (1-tailed)	.479	.000	.000	.	.000	
	N	46	46	46	1	46	46

**. Correlation is significant at the 0.01 level (1-tailed).

a. Cannot be computed because at least one of the variables is constant.

Appendix 34

Correlation between Bias Score and Academic Performance

Postgraduate Students

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.175*	-.122	.009	-.126	-.029
	Sig. (1-tailed)		.026	.095	.487	.083	.386
	N	127	123	117	14	123	106
Overall Mark	Pearson Correlation	-.175*	1.000	.661**	.034	.881**	.681**
	Sig. (1-tailed)	.026		.000	.454	.000	.000
	N	123	123	117	14	123	106
Exam Mark	Pearson Correlation	-.122	.661**	1.000	.168	.529**	.262**
	Sig. (1-tailed)	.095	.000		.283	.000	.004
	N	117	117	117	14	117	102
Presentation Mark	Pearson Correlation	.009	.034	.168	1.000	.058	-.095
	Sig. (1-tailed)	.487	.454	.283		.423	.390
	N	14	14	14	14	14	11
Coursework Mark	Pearson Correlation	-.126	.881**	.529**	.058	1.000	.570**
	Sig. (1-tailed)	.083	.000	.000	.423		.000
	N	123	123	117	14	123	106
Dissertation Mark	Pearson Correlation	-.029	.681**	.262**	-.095	.570**	1.000
	Sig. (1-tailed)	.386	.000	.004	.390	.000	
	N	106	106	102	11	106	106

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 35

Correlation between Bias Score and Academic Performance

Postgraduate Students – Direct Entrants

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.226*	-.192*	-.053	-.241*	-.075
	Sig. (1-tailed)		.022	.049	.438	.016	.277
	N	80	79	75	11	79	64
Overall Mark	Pearson Correlation	-.226*	1.000	.684**	.066	.917**	.692**
	Sig. (1-tailed)	.022		.000	.424	.000	.000
	N	79	79	75	11	79	64
Exam Mark	Pearson Correlation	-.192*	.684**	1.000	.251	.510**	.180
	Sig. (1-tailed)	.049	.000		.229	.000	.081
	N	75	75	75	11	75	62
Presentation Mark	Pearson Correlation	-.053	.066	.251	1.000	.045	-.085
	Sig. (1-tailed)	.438	.424	.229		.447	.414
	N	11	11	11	11	11	9
Coursework Mark	Pearson Correlation	-.241*	.917**	.510**	.045	1.000	.554**
	Sig. (1-tailed)	.016	.000	.000	.447		.000
	N	79	79	75	11	79	64
Dissertation Mark	Pearson Correlation	-.075	.692**	.180	-.085	.554**	1.000
	Sig. (1-tailed)	.277	.000	.081	.414	.000	
	N	64	64	62	9	64	64

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 36

Correlation between Bias Score and Academic Performance

Postgraduate Continuing Students

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.099	.059	.921	.082	.029
	Sig. (1-tailed)		.261	.355	.128	.299	.427
	N	47	44	42	3	44	42
Overall Mark	Pearson Correlation	-.099	1.000	.645**	.740	.806**	.672**
	Sig. (1-tailed)	.261		.000	.235	.000	.000
	N	44	44	42	3	44	42
Exam Mark	Pearson Correlation	.059	.645**	1.000	-.579	.623**	.388**
	Sig. (1-tailed)	.355	.000		.303	.000	.007
	N	42	42	42	3	42	40
Presentation Mark	Pearson Correlation	.921	.740	-.579	1.000	.805	1.000**
	Sig. (1-tailed)	.128	.235	.303		.202	.
	N	3	3	3	3	3	2
Coursework Mark	Pearson Correlation	.082	.806**	.623**	.805	1.000	.595**
	Sig. (1-tailed)	.299	.000	.000	.202		.000
	N	44	44	42	3	44	42
Dissertation Mark	Pearson Correlation	.029	.672**	.388**	1.000**	.595**	1.000
	Sig. (1-tailed)	.427	.000	.007	.	.000	
	N	42	42	40	2	42	42

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 37

Test for Differences in Bias Score by Entry Status

All Students

Group Statistics

	Entry Status	N	Mean	Std. Deviation	Std. Error Mean
Bias Score	Direct Entrants	250	.0997	.14526	.00919
	Continuing Students	258	.0776	.14941	.00930

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Bias Score	Equal variances assumed	.225	.635	1.688	506	.092	.02208	.01308	-.00361	.04778
	Equal variances not assumed			1.689	505.994	.092	.02208	.01307	-.00360	.04777

Appendix 38

Correlation between Bias Score and Academic Performance

All Direct Entrants

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.190**	-.171**	.022	-.199**	-.077
	Sig. (1-tailed)		.002	.004	.433	.001	.239
	N	250	240	236	62	240	86
Overall Mark	Pearson Correlation	-.190**	1.000	.887**	.169	.864**	.711**
	Sig. (1-tailed)	.002		.000	.095	.000	.000
	N	240	240	236	62	240	86
Exam Mark	Pearson Correlation	-.171**	.887**	1.000	.129	.722**	.335**
	Sig. (1-tailed)	.004	.000		.159	.000	.001
	N	236	236	236	62	236	84
Presentation Mark	Pearson Correlation	.022	.169	.129	1.000	.246*	-.090
	Sig. (1-tailed)	.433	.095	.159		.027	.402
	N	62	62	62	62	62	10
Coursework Mark	Pearson Correlation	-.199**	.864**	.722**	.246*	1.000	.619**
	Sig. (1-tailed)	.001	.000	.000	.027		.000
	N	240	240	236	62	240	86
Dissertation Mark	Pearson Correlation	-.077	.711**	.335**	-.090	.619**	1.000
	Sig. (1-tailed)	.239	.000	.001	.402	.000	
	N	86	86	84	10	86	86

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

Appendix 39

Correlation between Bias Score and Academic Performance

All Continuing Students

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.023	-.021	.206	-.067	.077
	Sig. (1-tailed)		.365	.378	.222	.157	.269
	N	258	229	227	16	227	66
Overall Mark	Pearson Correlation	-.023	1.000	.837**	.373	.843**	.773**
	Sig. (1-tailed)	.365		.000	.077	.000	.000
	N	229	229	227	16	227	66
Exam Mark	Pearson Correlation	-.021	.837**	1.000	.373	.626**	.502**
	Sig. (1-tailed)	.378	.000		.078	.000	.000
	N	227	227	227	16	225	64
Presentation Mark	Pearson Correlation	.206	.373	.373	1.000	.251	1.000**
	Sig. (1-tailed)	.222	.077	.078		.174	.
	N	16	16	16	16	16	2
Coursework Mark	Pearson Correlation	-.067	.843**	.626**	.251	1.000	.742**
	Sig. (1-tailed)	.157	.000	.000	.174		.000
	N	227	227	225	16	227	66
Dissertation Mark	Pearson Correlation	.077	.773**	.502**	1.000**	.742**	1.000
	Sig. (1-tailed)	.269	.000	.000	.	.000	
	N	66	66	64	2	66	66

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 40

Correlation between Bias Score and Academic Performance

Aged 18-21 Years

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.074	-.061	.061	-.070	-.219
	Sig. (1-tailed)		.115	.159	.328	.129	.237
	N	297	268	268	56	266	13
Overall Mark	Pearson Correlation	-.074	1.000	.906**	.249*	.821**	.697**
	Sig. (1-tailed)	.115		.000	.032	.000	.004
	N	268	268	268	56	266	13
Exam Mark	Pearson Correlation	-.061	.906**	1.000	.114	.658**	.548*
	Sig. (1-tailed)	.159	.000		.201	.000	.026
	N	268	268	268	56	266	13
Presentation Mark	Pearson Correlation	.061	.249*	.114	1.000	.347**	.[a]
	Sig. (1-tailed)	.328	.032	.201		.004	.
	N	56	56	56	56	56	0
Coursework Mark	Pearson Correlation	-.070	.821**	.658**	.347**	1.000	.752**
	Sig. (1-tailed)	.129	.000	.000	.004		.002
	N	266	266	266	56	266	13
Dissertation Mark	Pearson Correlation	-.219	.697**	.548*	.[a]	.752**	1.000
	Sig. (1-tailed)	.237	.004	.026	.	.002	
	N	13	13	13	0	13	13

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

a. Cannot be computed because at least one of the variables is constant.

Appendix 41

Correlation between Bias Score and Academic Performance

Aged 22+ Years

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.110	-.078	.059	-.153*	.001
	Sig. (1-tailed)		.062	.142	.396	.016	.496
	N	207	197	191	22	197	138
Overall Mark	Pearson Correlation	-.110	1.000	.782**	.033	.870**	.745**
	Sig. (1-tailed)	.062		.000	.442	.000	.000
	N	197	197	191	22	197	138
Exam Mark	Pearson Correlation	-.078	.782**	1.000	.162	.626**	.381**
	Sig. (1-tailed)	.142	.000		.236	.000	.000
	N	191	191	191	22	191	134
Presentation Mark	Pearson Correlation	.059	.033	.162	1.000	.014	-.105
	Sig. (1-tailed)	.396	.442	.236		.475	.372
	N	22	22	22	22	22	12
Coursework Mark	Pearson Correlation	-.153*	.870**	.626**	.014	1.000	.673**
	Sig. (1-tailed)	.016	.000	.000	.475		.000
	N	197	197	191	22	197	138
Dissertation Mark	Pearson Correlation	.001	.745**	.381**	-.105	.673**	1.000
	Sig. (1-tailed)	.496	.000	.000	.372	.000	
	N	138	138	134	12	138	138

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 42

Correlation between Bias Score and Academic Performance

Males

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.117*	-.099	.068	-.120*	.036
	Sig. (1-tailed)		.030	.055	.353	.026	.379
	N	281	261	259	33	260	77
Overall Mark	Pearson Correlation	-.117*	1.000	.854**	.415**	.861**	.731**
	Sig. (1-tailed)	.030		.000	.008	.000	.000
	N	261	261	259	33	260	77
Exam Mark	Pearson Correlation	-.099	.854**	1.000	.221	.672**	.356**
	Sig. (1-tailed)	.055	.000		.108	.000	.001
	N	259	259	259	33	258	76
Presentation Mark	Pearson Correlation	.068	.415**	.221	1.000	.514**	1.000**
	Sig. (1-tailed)	.353	.008	.108		.001	.
	N	33	33	33	33	33	2
Coursework Mark	Pearson Correlation	-.120*	.861**	.672**	.514**	1.000	.656**
	Sig. (1-tailed)	.026	.000	.000	.001		.000
	N	260	260	258	33	260	77
Dissertation Mark	Pearson Correlation	.036	.731**	.356**	1.000**	.656**	1.000
	Sig. (1-tailed)	.379	.000	.001	.	.000	
	N	77	77	76	2	77	77

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 43

Correlation between Bias Score and Academic Performance

Females

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.113	-.101	.042	-.171**	-.058
	Sig. (1-tailed)		.054	.078	.394	.007	.312
	N	224	205	201	44	204	75
Overall Mark	Pearson Correlation	-.113	1.000	.885**	.052	.847**	.747**
	Sig. (1-tailed)	.054		.000	.369	.000	.000
	N	205	205	201	44	204	75
Exam Mark	Pearson Correlation	-.101	.885**	1.000	.094	.699**	.458**
	Sig. (1-tailed)	.078	.000		.272	.000	.000
	N	201	201	201	44	200	72
Presentation Mark	Pearson Correlation	.042	.052	.094	1.000	.064	-.158
	Sig. (1-tailed)	.394	.369	.272		.340	.331
	N	44	44	44	44	44	10
Coursework Mark	Pearson Correlation	-.171**	.847**	.699**	.064	1.000	.710**
	Sig. (1-tailed)	.007	.000	.000	.340		.000
	N	204	204	200	44	204	75
Dissertation Mark	Pearson Correlation	-.058	.747**	.458**	-.158	.710**	1.000
	Sig. (1-tailed)	.312	.000	.000	.331	.000	
	N	75	75	72	10	75	75

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix 44

Correlation between Bias Score and Academic Performance

UK Students

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	.013	.030	.126	.048	.119
	Sig. (1-tailed)		.422	.320	.216	.229	.281
	N	269	243	243	41	241	26
Overall Mark	Pearson Correlation	.013	1.000	.871**	.283*	.811**	.754**
	Sig. (1-tailed)	.422		.000	.036	.000	.000
	N	243	243	243	41	241	26
Exam Mark	Pearson Correlation	.030	.871**	1.000	.126	.625**	.392*
	Sig. (1-tailed)	.320	.000		.216	.000	.024
	N	243	243	243	41	241	26
Presentation Mark	Pearson Correlation	.126	.283*	.126	1.000	.414**	.[a]
	Sig. (1-tailed)	.216	.036	.216		.004	.
	N	41	41	41	41	41	0
Coursework Mark	Pearson Correlation	.048	.811**	.625**	.414**	1.000	.794**
	Sig. (1-tailed)	.229	.000	.000	.004		.000
	N	241	241	241	41	241	26
Dissertation Mark	Pearson Correlation	.119	.754**	.392*	.[a]	.794**	1.000
	Sig. (1-tailed)	.281	.000	.024	.	.000	
	N	26	26	26	0	26	26

***. Correlation is significant at the 0.01 level (1-tailed).*

**. Correlation is significant at the 0.05 level (1-tailed).*

a. Cannot be computed because at least one of the variables is constant.

Appendix 45

Correlation between Bias Score and Academic Performance

Chinese Students

Correlations

		Bias Score	Overall Mark	Exam Mark	Presentation Mark	Coursework Mark	Dissertation Mark
Bias Score	Pearson Correlation	1.000	-.008	-.082	-.595*	-.010	.167
	Sig. (1-tailed)		.470	.217	.035	.461	.108
	N	102	98	94	10	98	57
Overall Mark	Pearson Correlation	-.008	1.000	.754**	-.487	.799**	.753**
	Sig. (1-tailed)	.470		.000	.077	.000	.000
	N	98	98	94	10	98	57
Exam Mark	Pearson Correlation	-.082	.754**	1.000	-.438	.522**	.192
	Sig. (1-tailed)	.217	.000		.103	.000	.084
	N	94	94	94	10	94	53
Presentation Mark	Pearson Correlation	-.595*	-.487	-.438	1.000	-.315	1.000**
	Sig. (1-tailed)	.035	.077	.103		.187	.
	N	10	10	10	10	10	2
Coursework Mark	Pearson Correlation	-.010	.799**	.522**	-.315	1.000	.666**
	Sig. (1-tailed)	.461	.000	.000	.187		.000
	N	98	98	94	10	98	57
Dissertation Mark	Pearson Correlation	.167	.753**	.192	1.000**	.666**	1.000
	Sig. (1-tailed)	.108	.000	.084	.	.000	
	N	57	57	53	2	57	57

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

List of References

Acker, D. and Duck, N.W. (2008) 'Cross-cultural overconfidence and biased self-attribution', *The Journal of Socio-Economics*, 37, pp. 1815-1824.

Ackerman, P.L., Beier, M.E. and Bowen, K.R. (2002) 'What we really know about our abilities and our knowledge', *Personality and Individual Differences*, 33(4), pp. 587–605.

Adams, J.K. and Adams, P.A. (1961) 'Realism of confidence judgements', *Psychological Review*, 68, pp. 33-45.

Alicke, M.D., Klotz, M.L., Breitenbecher, D.L., Yurak, T.J. and Vredenburg, D.S. (1995) 'Personal contact, individuation, and the better-than-average effect', *Journal of Personality and Social Psychology*, 68 (5), pp.804-25.

Allwood, C.M. and Granhag, P.A. (1999) 'Feelings of confidence and the realism of confidence judgements in everyday life' in, Juslin, P. and Montgomery, H. (eds.). *Judgement and Decision Making: Neo-Brunswikian and Process-Tracing Approaches*. New Jersey, Lawrence Erlbaum Associates.

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. and Wittrock, M.C. (eds.) (2001) *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, Longman.

Aram, J.D. and Salipante, P.F. (2003) 'Bridging scholarship in management: Epistemological reflections', *British Journal of Management*, 16, pp.189-205.

Arkes, H.R., Christensen, C., Lai, C. and Blumer, C. (1987) 'Two methods of reducing overconfidence', *Organizational Behavior and Human Decision Processes*, 39, pp. 133-144.

Azevedo, R., (2009) 'Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: a discussion', *Metacognition Learning*, 4(1), pp. 87–95.

Babad, E. (1987) 'Wishful thinking and objectivity among sports fans', *Social Behavior: an International Journal of Applied Social Psychology*, 4(2), pp. 231-240.

Baker, T.L. (1999) *Doing Social Research* (3rd. Edn.), McGraw Hill.

Bar-Tal, Y., Sarid, A., and Kishon-Rabin, L. (2001) 'A Test of Overconfidence Phenomenon Using Audio Signals', *Journal of General Psychology*, 128 (1), pp. 76-80.

Baranski, J.V. and Petrusic, W.M. (1999) 'Realism of confidence in sensory discrimination' *Perception and Psychophysics*, 61(7), pp. 1369-1383.

Barber, B.M and Odean, T. (2001) 'Boys Will Be Boys: Gender, Overconfidence and common Stock Investment', *Quarterly Journal of Economics*, 116 (1).

BBC (2003) 'Rum Remark Wins Rumsfeld an Award' [Online], Available at http://news.bbc.co.uk/1/hi/world/americas/3254852.stm (Accessed 31 July 2008).

Bell, J. (1993) *Doing Your Research Project*, Buckingham: Open University Press.

Bem, D.J. and Lord, C. (1979) 'Template matching: a proposal for probing the ecological validity of experimental settings in social psychology' *Journal of Personality and Social Psychology*, 37 (6), pp. 833-46.

Beyer, S. (1990) 'Gender differences in the accuracy of self-evaluations of performance', *Journal of Personality and Social Psychology*, 59(5), pp. 960-970.

Bjorkman, M. (1994) 'Internal cue theory: calibration and resolution of confidence in general knowledge' *Organizational Behavior and Human Decision Processes*, 58, pp. 386-405.

Black, T.R. (1993) *Evaluating Social Science Research: An Introduction*. London: Sage.

Bloom, B.S. (Ed). (1956) *Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook 1: Cognitive Domain*. London: David McKay.

Boud, D.J. (1995) 'How Can Self-Assessment Strategies be Designed?' in Boud, D.J. (Ed.) *Enhancing Learning Through Self-Assessment*. London: Kogan Page.

Boud, D.J. and Falchikov, N. (1995) 'What Does Research Tell Us About Self-Assessment?' in Boud, D.J. (Ed.) *Enhancing Learning Through Self-Assessment*. London: Kogan Page.

Boud, D.J. and Falchikov, N. (1989) 'Quantitative studies of student self-assessment in higher education: a critical analysis of findings' *Higher Education*, 18. pp. 529-549.

Braun, P.A. and Yaniv, I. (1992) 'A case study of experiment judgement: economists' probabilities versus base-rate model forecasts', *Journal of Behavioral Decision Making*, 5, pp. 217-231.

Bromily, P. and Curley, S.P. (1992) 'Individual Differences' in Yates, F.J. (Ed.), *Risk Taking Behaviour* Chichester: Wiley and Sons.

Brown, J.D. and Gallagher, F.M. (1992) 'Coming to terms with failure: private self-enhancement and public self-effacement' *Journal of Experimental Social Psychology*, 28 pp. 3-22.

Bryman, A. (2008) *Social Research Methods*, (3rd. Edn.). Oxford: Oxford University Press.

Burson, K. A., Larrick, R. P., and Klayman, J. (2006) 'Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons' *Journal of Personality and Social Psychology*, 90(1), pp. 60–77.

Camerer, C.F. and Hogarth, R.M. (1999) 'The effects of financial incentives in experiments: a review and capital-labor production framework', *Journal of Risk and Uncertainty*, 19 (1-3) pp. 7-42.

Camerer, C. and Lovallo, D. (1999) 'Overconfidence and excess Entry: An Experimental Approach', *American Economic Review*, 89 (1) pp. 306-318.

Charba, J.P. and Klein, W.H. (1980) 'Skill in precipitation forecasting in the national weather service', *Bulletin of the American Meteorological Society*, 61(12), pp. 1546-1555.

Chen, C., Lee, S. and Stevenson, H.W. (1995) 'Response style and cross-cultural comparisons of rating scales among east Asian and North American students', *Psychological Science*, 6(3) pp. 170-175.

Christensen-Szalanski, J.J.J. and Bushyhead, J.B. (1981) 'Physicians' use of probabilistic information in a real clinical setting', *Journal of Experimental Psychology: Human Perception and Performance*, 7(4), pp. 928-935.

Clarebout, G., Elen, J. and Onghena, P. (2006) 'In search of the reliability of a Flemish version of the Knowledge Monitoring Assessment Test', *Metacognition Learning*, 1(1) pp. 137–147.

Cohen, L., Manion, L. and Morrison, K. (2000) *Research Methods in Education*, (5th Edn.) Abingdon: Routledge Falmer.

Copleston, F. (1963) *A History of Modern Philosophy: Volume 4, Modern Philosophy: Descartes to Leibniz.* New York: Doubleday.

Crittenden, V. and Woodside, A.G. (2007) 'Building skills in thinking: toward a pedagogy in metathinking', *Journal of Education for Business,* Sept/Oct, pp. 37-43.

Crotty, M. (1998) *The Foundations of Social Research: Meaning and Perspective in the Research Process.* London: Sage.

Culpepper, R.A., Zhao, L. and Lowery, C. (2002) 'Survey response bias among Chinese managers', *Academy of Management Proceedings* 2002 IM:J1.

Czaja, R. and Blair, J. (1996) *Designing Surveys: A Guide to Decisions and Procedures.* London: Sage.

Davidoff, F. (1995) 'Confidence testing – how to answer a meta-question.' *American College of Physicians Observer [Online],* Available at, http://*www.acponline.org/journals/news/feb95/conftest.htm (Accessed 24 February, 2009).*

Davis, D. (2005) *Business Research for Decision Making,* (6th Edn.), London: Thomson Brooks/Cole.

Deaux, K. and Farris, E., (1977) 'Attributing causes for one's own performance: the effects of sex, norms and outcome', *Journal of Research in Personality*, 11 pp. 59-72.

Denscombe, M. (2003) *The Good Research Guide: for Small-scale Social Research Projects.* Maidenhead: Open University Press.

Dunlosky, J. and Hertzog, C. (2000) 'Updating knowledge about encoding strategies: a componential analysis of learning about strategy effectiveness from task experience', *Psychology and Aging*, 15(3), pp. 462-474.

Dunning, D., Griffin, D.W., Milojkovic, J.D. and Ross, L. (1990) 'The overconfidence effect in social prediction', *Journal of Personality and Social Psychology*, 58 (4), pp. 568-81.

Dunning, D., Johnson, K., Ehrlinger, J., and Kruger, J. (2003) 'Why people fail to recognize their own incompetence', *Current Directions in Psychological Science*, 12(3), pp. 83–87.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D. and Kruger, J. (2008) 'Why the unskilled are unaware: further explorations of (absent) self-insight among the incompetent', *Organizational Behavior and Human Decision Processes,* 105, pp. 98-121.

Eilon, S. (1974) 'Seven faces of research', *Omega*, 2(1), pp. 1-9.

Ekman, P. and O'Sullivan, M. (1991) 'Who can catch a liar?' *American Psychologist*, 46(9), pp. 913-920.

Erev, I. and Cohen, B.L. (1990) 'Verbal versus numerical probabilities: efficiency, biases, and the preference paradox', *Organizational Behavior and Human Decision Processes,* 45, pp.1-18.

Everson, H.T. and Tobias, S. (1998) 'The ability to estimate knowledge and performance in college: a metacognitive analysis', *Instructional Science,* 26 pp. 65–79.

Falchikov, N. and Boud, D. (1989) 'Student self-assessment in higher education: a meta-analysis' *Review of Educational Research*, 59(4) pp. 395-430.

Farrell, G. and Leung, Y.K. (2004) 'Innovative online assessment using confidence measurement', *Education and Information Technologies*, 9(1) pp. 5–19.

Field, A. and Hole, G. (2003) *How to Design and Report Experiments*. London: Sage.

Fischhoff, B. (1994) 'What forecasts (seem to) mean', *International Journal of Forecasting*, 10(3) pp. 387-403.

Fischhoff, B. (1992) 'Risk taking: a developmental perspective', in Yates, F.J. (ed.), *Risk Taking Behaviour* Chichester: Wiley and Sons.

Fischhoff, B. (1982) 'Debiasing' in, Kahneman, D., Slovic, P. and Tversky, A. (eds.) *Judgement under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Fischhoff, B. and MacGregor, (1982), 'Subjective confidence in forecasts' *Journal of Forecasting*, 1, pp. 155-172

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1977) 'Knowing with certainty: the appropriateness of extreme confidence', *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), pp. 552-564.

Fisher, C. (2004) *Researching and Writing a Dissertation: A Guidebook for Business Students* 2nd. Edn. Harlow: Pearson.

Fitzgerald, J.T., White, C.W. and Gruppen, L.D. (2003) 'A longitudinal study of self-assessment accuracy' *Medical Education*, 37 pp. 645-649.

Flannelly, L.T. (2001) 'Using feedback to reduce students' judgment bias on test questions', *Journal of Nursing Education* 40(1) pp. 10-16.

Flannelly, L.T. and Flannelly, K.J. (2000) 'Reducing people's judgement bias about their level of knowledge', *The Psychological Record*, 50(3), pp. 587-600.

Flavell, J. H. (1979) 'Metacognition and cognitive monitoring: a new area of cognitive-developmental enquiry', *American Psychologist*, 34(10), pp. 906-911.

Gersten, R., Vaughn, S., Deshler, D. and Schiller, E. (1997) 'What we know about using research findings: Implications for improving special education practice', *Journal of Learning Disabilities*, 30(5) pp.466-476.

Gigerenzer, G. (1991) 'How to make cognitive illusions disappear: beyond heuristics and biases', *European Review of Social Psychology, 2, pp* .83-115.

Gilfoyle, J. (2000) Behavioural Finance 101, *Benefits Canada*, 24 (12) pp. 76-78.

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

Great Britain, Data Protection Act 1998, Elizabeth II, Chapter 29 (1998) [Online], Available at http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_1 (Accessed 29 July 2009).

Gredler, M. E. (2004) 'Games and simulations and their relationships to learning', in Jonassen, D.H. and Harris, P. (eds.), *Handbook of research on educational communications and technology*. pp. 571–582. Mahwah, NJ: Erlbaum.

Grimes, P.W. (2002) 'The Overconfident principles of economics students: an examination of a metacognitive skill' *Journal of Economic Education*, Winter pp.15-30.

Hacker, D., (1998) 'Definitions and Empirical Foundations' in, Hacker, D., Dunlosky, J. and Graesser, A.C. (Eds.) *Metacognition in Educational Theory and Practice*. London: Lawrence Erlbaum Associates.

Hacker, D.J., Bol, L., and Bahbahani (2008) 'Explaining calibration accuracy in classroom contexts: the effect of incentives, reflection, and explanatory style', *Metacognition Learning*, 3, pp. 101-121.

Hacker, D.J., Bol, L., Horgan, D.D. and Rakow, E.A. (2000) 'Test prediction and performance in a classroom context', *Journal of Educational Psychology*, 92(1) pp. 160-170.

Halfpenny, P. (1982) *Positivism and Sociology: Explaining Social Life*. London: Unwin.

Hammond, J.S., Keeney, R.L. and Raiffa, H. (1998) 'The hidden traps in decision making', *Harvard Business Review*, 76(5) pp.47-54.

Hansen, J. D., and Dexter, L., (1997) 'Quality multiple-choice test questions: Item-writing guidelines and an Analysis of Auditing Testbanks' *Journal of Education for Business* 73(2) pp.94-97.

Harvey, N. (1997) 'Confidence in judgement', *Trends in Cognitive Sciences* 1(2) pp. 78-82.

Hemsley-Brown, J.V. and Sharp, C. (2003) 'The use of research to improve professional practice: A systematic review of the literature' *Oxford Review of Education*, 29(4), pp.449-470.

Hillage, J., Pearson, R., Anderson, A. and Tamkin, P. (1998) '*Excellence in Research in Schools'* DfEE Research Report 74, London, DfEE.

Hinton, P.R., Brownlow, C., McMurray, I., and Cozens, B. (2004) *SPSS Explained*. Hove: Routledge.

Huberman, M. (1990) 'Linkage between researchers and practitioners: A qualitative study', *American Educational Research Journal*, 27(2) pp. 363–391.

Hughes, J.A. and Sharrock, W.W. (2007) *Theory and Methods in Sociology: An Introduction to Sociological Thinking and Practice*. Basingstoke: Palgrave Macmillan.

Isaacson, R.M. and Fujita, F. (2006) 'Metacognitive knowledge monitoring and self-regulated learning: academic success and reflections on learning' *Journal of the Scholarship of Teaching and Learning*, 6(1), pp. 39-55.

James, K. and Vinnicombe, S. (2002) 'Acknowledging the Individual in the Researcher' in, Partington, D. Ed. (2002) *Essential Skills for Management Research*. London: Sage.

Jankowicz, A.D. (1995) *Business Research Projects* (2nd. edn.), London: Chapman and Hall.

Jehng, J.C., Johnson, S.D. and Anderson, R.C. (1993) 'Schooling and students' epistemological beliefs about learning', *Contemporary Educational Psychology*, 18 (1), pp. 23-35.

Johnson, P. and Duberley, J. (2000) *Understanding Management Research*. London: Sage.

Juslin, P. (1994) 'The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items', *Organisational Behaviour and Human Decision Processes* 57, pp. 226-246.

Kemp and Holmwood (2003) 'Realism, regularity and social explanation', *Journal for the Theory of Social Behaviour,* 33(2), pp. 165-187.

Kennedy, E.J., Lawton, L. and Plumlee, E.L. (2002) 'Blissful ignorance: the problem of unrecognized incompetence and academic performance' *Journal of Marketing Education*, 24 (3) pp. 243-252.

Keren, G. (1997) 'On the calibration of probability judgements: some critical comments and alternative perspectives', *Journal of Behavioural Decision Making* 10, pp. 269-278.

Keren, G. (1991) 'Calibration and probability judgements: conceptual and methodological issues' *Acta Psychologica* 77 pp. 217-273.

Keren, G. (1987) 'Facing uncertainty in the game of bridge: a calibration study.' *Organizational Behavior and Human Decision Processes,* 39 pp. 98-114.

Klayman, J., Soll, J.B., Gonzalez-Vallejo, C. and Barlas, S. (1999) 'Overconfidence: it depends on how, what, and whom you ask' *Organizational Behavior and Human Decision Processes,* 79(3), pp. 216-247.

Klenowski, V. (1995) 'Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England' *Assessment in Education*, 2(2), pp.145-163.

Koku, P.S. and Qureshi, A.A. (2004) 'Overconfidence and the performance of business students on examinations' *Journal of Accounting for Business*, March/April, pp. 217-224.

Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980) 'Reasons for confidence', *Journal of Experimental Psychology: Human Learning and Memory,* 6(2), pp. 107-118.

Krathwohl, D.R. (2002) 'A revision of Bloom's taxonomy: an overview', *Theory into Practice,* 41(4), pp. 212-218.

Kruger, J. and Dunning, D. (1999) 'Unskilled and unaware of it: how difficulties in recognising one's own incompetence lead to inflated self-assessments', *Journal of Personality and Social Psychology,* 77(6), pp. 1121-34.

Larres, P.M., Ballantine, J.A., and Whittington, M. (2003) 'Evaluating the validity of self-assessment: measuring computer literacy among entry-level undergraduates within accounting degree programmes at two UK universities', *Accounting Education*, 12(2), pp. 97-112.

Lichtenstein, S., and Fischhoff, B. (1980) 'Training for calibration', *Organizational Behavior and Human Performance*, 26, pp. 149-171.

Lichtenstein, S., and Fischhoff, B. (1977) 'Do those who know more also know more about how much they know?', *Organizational Behavior and Human Performance,* 20, pp. 159-183.

Lichtenstein, S., Fischhoff, B. and Phillips, L.D. (1982) 'Calibration of probabilities: the state of the art to 1980' in, Kahneman, D., Slovic, P. and Tversky, A. (eds.) *Judgement under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Loftus, E.F. and Wagenaar, W.A. (1988) 'Lawyers predictions of success', *Jurimetrics Journal* 28, pp. 437-453.

Lories, G., Dardenne, B. and Yzerbyt, V.Y. (1998) 'From social cognition to metacognition' in, Yzerbyt, V.Y., Lories, G. and Dardenne, B. (eds.), *Metacognition: Cognitive and Social Dimensions*. London: Sage.

Lundeberg, M.A., Fox, P.W. and Punccohar, J. (1994) 'Highly confident but wrong: gender differences and similarities in confidence judgments', *Journal of Educational Psychology*, 86(1), pp. 114-121.

Maykut, P. and Morehouse, R. (1994) *Beginning Qualitative Research: A Philosophical and Practical Guide*. London: The Falmer Press.

McGregor, S. (2004) ‘Learning with Donald Rumsfeld — flexible learning: the relevance and resonance of multi-professional learning in primary care’, *British Journal of General Practice*, September (54) pp. 722–723.

Meehan, A.M. and Overton, W.F. (1986) ‘Gender differences in expectancies for success and performance on Piagetian spatial tasks’, *Merrill-Palmer Quarterly*, 32(4) pp. 427-441.

Montalvo, F.T. and Torres, M.C.G. (2004) ‘Self regulated learning: current and future directions’, *Electronic Journal of Research in Educational Psychology*, 2(1), pp. 1-34.

Mowl, G. and Pain, R. (1995) ‘Using self and peer assessment to improve students' essay writing: a case study from geography’, *Innovations in Education and Teaching International*, 32(4) pp. 324-335.

Muijs, D. (2004) *Doing Quantitative Research in Education with SPSS*, London: Sage.

Murphy, A.H. and Winkler, R.L. (1984) ‘Probability forecasting in meteorology’, *Journal of the American Statistical Association*, 79, pp. 489-500.

Nash, R. (2005) ‘Explanation and quantification in educational research: the arguments of critical and scientific realism’, *British Educational Research Journal*, 31(2), pp. 185-204.

Nash, R. (1999) ‘Realism in the sociology of education: ‘explaining’ social differences in attainment’ *British Journal of Sociology of Education*, 20 pp. 107-125.
Nelson, T.O. (1996) ‘Consciousness and metacognition’, *American Psychologist*, 51(2), pp. 102-116.

Nicholson, N. and Willman, P. (2001) ‘Folly, fantasy and roguery: a social psychology of finance risk disasters’, in Pickford, J. (ed.) *Mastering Risk Volume 1: Concepts*. Harlow: Financial Times Prentice Hall.

Nietfield, J.L., Cao, L. and Osborne, J.W. (2006) 'The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy', *Metacognition Learning*, 1, pp. 159–179.

Oskamp, S. (1965) 'Overconfidence in case study judgements', *The Journal of Consulting Psychology*, 29, pp. 261-265.

Oskamp, S. (1962) 'The relationship of clinical experience and training methods to several criteria of clinical prediction' *Psychological Monographs*, 76, No.547.

Pallier, G. (2003) 'Gender differences in the self-assessment of accuracy on cognitive tasks', *Sex Roles*, 48(5/6) pp. 265-276.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G. Stankov, L. and Roberts, R.D. (2002) 'The role of individual differences in the accuracy of confidence judgements', *The Journal of General Psychology* 129(3), pp. 257-299.

Paris, S.G. and Paris, A.H. (2001) 'Classroom applications of research on self-regulated learning', *Educational Psychologist*, *36*(2), pp. 89–101.

Paris, S.G. and Winograd, P. (2003) 'The role of self-regulated learning in contextual teaching: Principles and practices for teacher preparation', *Preparing Teachers to Use Contextual Teaching and Learning Strategies To Improve Student Success In and Beyond School* (Information Series No. 376) Columbus, Ohio: ERIC Clearinghouse on Adult, Career, and Vocational Education; Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education.

Paris, S.G. and Winograd, P. (1990) 'How metacognition can promote academic learning and instruction' in Jones, B.F and Idol, L. (eds.), *Dimensions of Thinking and Cognitive Instruction*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Pieschl, S. (2009) 'Metacognitive calibration—an extended conceptualization and potential applications', *Metacognition Learning*, 4(1) pp. 3–31.

Pintrich, P.R. (2002) ‘The role of metacognitive knowledge in learning, teaching, and assessing’, *Theory into Practice*, 41(4) pp. 219-225.

Plous, S. (1993) *The Psychology of Judgement and Decision Making*, New York: McGraw-Hill.

Poulton, E.C. (1994) *Behavioral Decision Theory: A New Approach*, Cambridge: Cambridge University Press.

Pressley, M. and Gaskins, I.W. (2006) ‘Metacognitively competent reading comprehension is constructively responsive reading: how can such reading be developed in students?’ *Metacognition Learning,* 1(1) pp. 99–113.

Pressley, M., Levi, J.R., Ghatala, E.S., and Ahmad, M. (1987) ‘Test monitoring in young grade school children’, *Journal of Experimental Child Psychology*, 43, pp. 96-111.

Pring, R. (2000) ‘The ‘false dualism’ of educational research’, *Journal of Philosophy of Education*, 34 (2), pp. 247-260.

Pulford, B.D. and Colman, A.M. (1997) ‘Overconfidence: feedback and item difficulty effects’, *Personality and Individual Differences*, 23(1), pp. 125-133.

Ramnarayan, S., Strohschneider, S., and Schaub, H. (1997),Trappings of expertise and the pursuit of failure’, *Simulation and Gaming*, 28(1), pp. 28–43.

Reeves, T. K. and Harper, D. G. (1981) *Surveys at work: a practitioner’s guide*. London: McGraw Hill.

Renner, C. H. and Renner M.J. (2001) ‘But I thought I knew that: using confidence estimation as a debiasing technique to improve classroom performance’, *Applied Cognitive Psychology* 15, pp. 23-32.

Robson, C. (2002) *Real World Research* (2nd. Edn.). Oxford: Blackwell.

Ronis, D.I. and Yates, J.F. (1987) 'Components of probability judgement accuracy: individual consistency and effects of subject matter and assessment method', *Organizational Behavior and Human Decision Processes*, 40, pp. 193-218.

Rosenthal, R. and Rosnow, R.L. (2008) *Essentials of Behavioral Research: Methods and Data Analysis*, (3rd Edn.). New York: McGraw Hill.

Rumsfeld, D. (2002) USA Department of Defense News Briefing Feb 12th, [Online], Available at http://www.telegraph.co.uk/money/main.jhtml?xml=/money/2006/09/19/ccinv19.xml, (Accessed 31 July 2008)

Russo, J.E. and Schoemaker, P.J.H. (1992) 'Managing overconfidence', *Sloan Management Review*, 33 (2), pp. 7-18.

Russo, J.E. and Schoemaker, P.J.H. (1989) *Decision Traps: Ten Barriers to Brilliant Decision Making and How to Overcome Them*. New York: Simon and Schuster.

Sarantakos, S. (2005) *Social Research* (3rd. Edn.), Basingstoke: Palgrave MacMillan.

Saunders, M., Lewis, P. and Thornhill, A. (2003) *Research Methods for Business Students.* (3rd. Edn.) Harlow: Pearson.

Schraw, G. (2009) 'A conceptual analysis of five measures of metacognitive monitoring', *Metacognition Learning*, 4 pp. 33–45.

Schraw, G. (1998) 'Promoting general metacognitive awareness', *Instructional Science* 26 pp. 113–125.

Schwab, D. P. (1985) 'Reviewing empirically based manuscripts: perspectives on progress', in Cummings, L.L. and Frost P. J., (eds.) *Publishing in the Organisation of Sciences,* Homewood Illinois: Irwin.

Scott, D. (2005) 'Critical realism and empirical research methods in education', *Journal of Philosophy of Education,* 39(4) pp. 633-646.

Shaughnessy, J.J. (1979) 'Confidence-judgement accuracy as a predictor of test performance', *Journal of Research in Personality,* 13, pp. 505-514.

Shen, C.E. and Pedulla, J.J. (2000) 'The relationship between students' achievement and their self-perception of competence and rigour of mathematics and science: a cross-national analysis', *Assessment in Education,* 7 (2) pp. 237-253.

Shuford, E. and Brown, T.A. (1975) 'Elicitation of personal probabilities and their assessment', *Instructional Science,* 4(2), pp. 137-188.

Sinkavich, F.J. (1995) 'Performance and metamemory: do students know what they don't know?' *Journal of Instructional Psychology*, 22(1), pp. 77-87.

Smith, D.J., Shields, W.E. and Washburn, D.A. (2003) 'The comparative psychology of uncertainty monitoring and metacognition', *Behavioral and Brain Sciences.,* 26, pp. 317–373.

Sparks, G.M. (1988) 'Teachers' attitudes toward change and subsequent improvements in classroom teaching', *Journal of Educational Psychology,* 80(1) pp. 111-117.

Stael von Holstein, C.A. (1972) 'Probabilistic forecasting: an experiment related to the stock market', *Organisational Behaviour and Human Performance,* 8. pp.139-158.

Sternberg, R.J. (1998) 'Metacognition, abilities, and developing expertise: What makes an expert student?' *Instructional Science* 26, pp. 127–140.

Stone, N. J. (2000) 'Exploring the relationship between calibration and self-regulated learning', *Educational Psychology Review*, 12(4), pp. 437–475.

Swartz, S.M. (2006) 'Acceptance and accuracy of multiple choice, confidence-level, and essay question formats for graduate students', *Journal of Education for Business*, March/April, pp. 215-220.

Tabachnick, B.G. and Fidel, L.S. (2001) *Using Multivariate Statistics* (4th Edn.), Needham Heights, MA: Pearson.

Taleb, N. N. (2004) *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets,* London: Penguin.

Ticehurst, G. W. and Veal, A. J. (1999) 'Business research methods: a managerial approach', Harlow, Addison Wesley Longman in, Minocha, S. (ed.) *Dissertation Preparation and Research Methods*, Harlow: Pearson.

Tobias, S., & Everson, H. (2002) '*Knowing What You Know and What You Don't: Further Research on Metacognitive Knowledge Monitoring*, College Board Report No. 2002-3, College Board, New York.

Tomassini, L.A., Solomon, I. Romney, M.B. and Krogstad, J.L. (1982) 'Calibration of auditors' probabilistic judgements: some empirical evidence', *Organizational Behavior and Human Performance*, 30, pp. 391-406.

Tversky, A. and Kahneman, D. (1982) 'Availability: a heuristic for judging frequency and probability', in Kahneman, D. Slovic, P. and Tversky, A. (eds.) *Judgement under uncertainty: Heuristics and biases*, Cambridge: Cambridge University Press.

Tversky, A. and Kahneman, D. (1974) 'Judgement under uncertainty: Heuristics and biases', *Science*, 185, pp. 1124-1131.

Tynjala, P., Helle, L. and Murtonen, M. (2001) 'Computer science and teacher students' beliefs concerning the nature of expertise in their field of study', presented at the 9th. European Conference on Learning and Instruction, Aug. 28 – Sep 1.

University of Northumbria (2009a) Ethics in Research – Policy Statement [Online], Available at

www.northumbria.ac.uk/researchandconsultancy/sa/ethgov/policies/?view=Standard (Accessed 29 July 2009).

University of Northumbria (2009b) Policy for Informed Consent in Research and Consultancy [Online], Available at

www.northumbria.ac.uk/researchandconsultancy/sa/ethgov/policies/?view=Standard (Accessed 29 July 2009).

Van Aken, J.E. (2005) 'Management research as a design science: Articulating the research products of mode 2 knowledge production in management', *British Journal of Management*, 16 pp.19-36.

Veenman, M.V.J., Van Hout-Wolters, B.H.A.M. and Afflerbach, P. (2006) 'Metacognition and learning: conceptual and methodological considerations', *Metacognition Learning* 1(1) pp. 3–14.

Wallace, M., and Wray, A. (2006) *Critical Reading and Writing for Postgraduates*, London: Sage.

Ward, M., Gruppen, L. and Regehr, G, (2002) 'Measuring self-assessment: current state of the art', *Advances in Health Science Education*, 7, pp. 63-80.

Weingardt, K. R., Leonesio, R. J., and Loftus, E. F. (1994) 'Viewing eyewitness research from a metacognitive perspective', in Metcalfe, J. and Shimamura A.P. (eds.), *Metacognition: Knowing about Knowing*, pp. 157–184, Cambridge: MIT.

Westley, D.N. (2008) 'Accuracy of student calibration on specific topics: very good students v's others', Proceedings of the Informing Science and I.T. Education Conference (1^{n} SITE).

Winch, C. (2001) 'Accountability and relevance in educational research', *Journal of Philosophy of Education,* 35(3), pp.443-459.

Wisker, G. (2001) *The Postgraduate Research Handbook: Succeed with Your MA, MPhil, EdD and PhD*. Basingstoke: Palgrave.

Worrall, L. (2008) 'Management research and management practice: is the relevance gap closing?', *International Journal of Management Concepts and Philosophy,* 3(1), pp.1-18

Wright, G. and Wishuda, A. (1982) 'Distribution of probability assessments for almanac and future event questions', *Scandinavian Journal of Psychology*, 23, pp. 219-224.

Yates, J.F., Lee, J.W., and Bush, J.G. (1997) 'General knowledge overconfidence: cross national variations, response style, and reality', *Organizational Behavior and Human Decision Processes,* 70(2), pp. 87-94.

Yates, J.F., Lee, J. and Shinotsuka, H. (1996) 'Beliefs about overconfidence, including its cross-national variation', *Organizational Behavior and Human Decision Processes,* 65(2), pp. 138-147.

Yates, J.F., Zhu, Y., Ronis, D.L., Wang, D., Shinotsuka, H. and Toda, M. (1989) 'Probability judgment accuracy: China, Japan, and the United States', *Organizational Behavior and Human Decision Processes,* (43) pp. 145-171.

Yzerbyt, V.Y., Lories, G. and Dardenne, B. (eds.) (1998) *Metacognition: Cognitive and Social Dimensions.* London: Sage.

Zakay, D. and Glicksohn, J. (1992) 'Overconfidence in a multiple-choice test and its relationship to achievement', *Psychological Record*, 42(4) pp.519-524.

Zenger, T. (1992) 'Why do employers only reward extreme performance? Examining the relationships among performance, pay, and turnover', *Administrative Science Quarterly*, 37(2) pp. 198-219.

Zeuli, J. (1994) 'How do teachers understand research when they read it'? *Teaching and Teacher Education*, 10, (1) pp. 39–55.

Bibliography

Argyrous, G. (2000) *Statistics for Social & Health Research*. London: Sage.

Bazerman, M.H. (2006) *Judgement in Managerial Decision Making*. Hoboken NJ: Wiley.

Beach, L.R. (1997) *The Psychology of Decision Making: People in Organisations*. London: Sage.

Bryman, A. and Cramer, D. (2005) *Quantitative Data Analysis with SPSS 12 and 13*. Hove: Routledge.

Hastie, R. and Dawes, R.M. (2001) *Rational Choice in an Uncertain World.* London: Sage.

Hussey, J. and Hussey, R. (1997) *Business Research: A Practical Guide for Undergraduate and Postgraduate Students.* Hampshire: MacMillan.

March, J.G. (1994) *A Primer on Decision Making: How Decisions Happen*. New York: MacMillan.

Mullins, G. and Kiley, M. (2002) 'It's a PhD, not a Nobel Prize: how research examiners assess research theses', *Studies in Higher Education,* 27(4) pp. 369-386.

Partington, D. (2002) *Essential Skills for Management Research*. London: Sage.