Exploring the problematic nature of GCSE examining in Economics and Business: Assessing troublesome knowledge, threshold concepts and learning.

> Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester

> > by

Andrew Ashwin B.Ed. (London) MBA (Hull) Department of Education University of Leicester

February 2015

Abstract

Title: Exploring the problematic nature of GCSE examining in Economics and Business: Assessing troublesome knowledge, threshold concepts and learning.

Andrew Ashwin

This thesis focuses upon assessment of learning at General Certificate of Secondary Education (GCSE) level. Approaches to learning, lecturers' conceptions of teaching, students' conceptions of learning, threshold concepts and troublesome knowledge have all been the focus of research at higher education but there has been limited work into the relevance of these fields to learning prior to higher education. This thesis surveys the research in higher education and applies some of the concepts to assessment, teaching and learning at lower levels of the education hierarchy. It looks at the extent to which students at GCSE level might be expected to begin the journey of thinking in the subject in the fields of economics and business. Teachers are a key influencer of assessment outcomes at GCSE level but their approach to teaching and their conception of learning may be influenced by the assessment framework in which they are operating.

Analysis of student responses to examination questions, the extent to which teachers at this level can agree on evidence of learning and what an assessment is designed to achieve and teachers' conceptions of learning will be studied at GCSE. The results of this research suggest that a reconceptualisation of the assessment objectives, which frame the specifications at this level and provide a focus for curriculum development, could influence the way students are taught and the way in which teaching and learning programmes are put together. Such a change could help to reduce the asymmetry between students and teachers and encourage teaching and learning which helps students to 'think in the subject' and champion deep approaches to learning.

Acknowledgements

I would like to acknowledge the efforts of students who embark on a journey of learning. As a former teacher it is easy to forget just what a challenge it is to begin such a journey. Deciding to do a PhD is a salient reminder of the challenges thousands of students undertake every year in schools. Their hard work and efforts, often in the face of considerable negative publicity, remain an inspiration. Thank you to all the students I have taught for teaching me so much and engaging my interest in assessment and economics and business.

Of course, any learning journey does not happen in isolation. The encouragement, belief but ultimately patience of my supervisors, Professor Steve Rayner and Professor Mary Fuller, has been a driving force to help me finally complete this part of my learning journey. Nicoleta Gaciu, a senior lecturer at Oxford Brookes University, also provided welcome support and guidance through elements of Phase II of the research.

Family and friends are also vital support mechanisms; my wife Sue and twin sons, Alex and Johnny, help to maintain a sense of balance and perspective in the midst of the submersion in academic study. My Mum never really understood education and qualifications but was always proud of whatever I did. I wish my Dad was alive to witness my graduation but I know he would have been equally as proud of me as Mum. To sister-in-law Cathy (the other Dr Ashwin in the family), you might have beaten me to it but, despite brother Michael's doubts, I did it in the end!

Finally, I would like to thank my colleagues at Cengage Learning EMEA. Cengage not only sponsored my research but offered unequivocal support and understanding throughout the years of my learning journey.

Table of Contents

1.	Intro	roduction 1		
	1.1	Background and context of the research1	1	
	1.2	The nature of the research problem1	3	
	1.3	Research Questions 1	6	
	1.4	Position of the Researcher1	6	
	1.5	Limitations of the Research1	9	
	1.6	Outline of Chapters 2	20	
	1.7	Summary of Chapter 2	21	
2.	Literature Review		23	
	2.1	Theories of Learning	23	
		2.1.1 Historical Perspectives on Learning	25	
		2.1.2Behavioural Theories of Learning	25	
		2.1.3Cognitive Theories of Learning	27	
2.2 The Purpose and Nature of Assessment and Ass		The Purpose and Nature of Assessment and Assessment Objectives		
			5	
		2.2.1 Definitions of Assessment 3	5	
		2.2.2 Reasons for Assessment	6	
		2.2.3 The Nature of Assessment	0	
		2.2.4 Validity and Reliability of Assessments	4	
		2.2.5Reliability 4	9	
		2.2.6Taxonomies and Educational Objectives	0	

		2.2.7The Structure of Observed Learning Outcome (SOLO)	68
	2.3	Approaches to Learning	75
		2.3.1Thinking in the Subject	78
	2.4	The Nature of Concepts	84
		2.4.1Theories of Concepts:	87
		2.4.2Concept Formation and Concept Acquisition	92
		2.4.3The Nature of Concepts	94
		2.4.4Conceptual Understanding	95
		2.4.5Conceptual Change	98
	2.5	Threshold Concepts and Troublesome Knowledge	102
		2.5.1 Implications of Threshold Concepts and Troublesome Knowledge	109
3.	Meth	hodology	112
	3.1	The Research Questions	112
	3.2	Philosophical Assumptions Underlying the Research	113
	3.3	Method and Methodology	117
	3.4	Position of the Researcher	120
	3.5	The Participants	123
		3.5.1Phase I	123
		3.5.2Phase II	124
		3.5.3Phase III	125
	3.6	Sources of Data	126
	3.7	Data Analysis	145
	3.8	Ethics	158
		F	2age 5

	3.9	Summary	
4.	Phas	se I: Results	
	4.1	Introduction161	
		4.1.1 The Assessment	
		4.1.2 Mark Schemes	
	4.2	The Results	
	4.3	Reflection on Coder Analysis: The Difference Between Content	
		Analysis and Text Analysis	
	4.4	Summary and Conclusion:170	
5.	Phas	se II: Results172	
	5.1	Introduction	
	5.2	Mark Schemes174	
		5.2.1 The Bloom Mark Scheme176	
		5.2.2 The SOLO Mark Scheme	
	5.3	Data Analysis Using Descriptive Statistics Assuming Parametric Data	
	5.4	Z- Scores	
		5.4.1 Tests for Normality194	
	5.5	Coefficient of Variation	
	5.6	Non-parametric Tests - Chi-square204	
	5.7	Summary and Conclusion210	
6.	Phas	Phase III: Results	
	6.1	Introduction	
	6.2	Results	
		6.2.1Question 1	

		6.2.2Question 2	217
		6.2.3Question 3	220
		6.2.4Question 4	224
		6.2.5Question 5	227
		6.2.6Question 6	231
		6.2.7Question 7	236
		6.2.8Question 8	238
		6.2.9Question 9	239
		6.2.10 Question 10	241
		6.2.11 Question 11	244
		6.2.12 Question 12	248
	6.3	Summary and Conclusion	251
7.	Disc	ussion, Conclusions and Recommendations	253
	7.1	Review of Outcomes of Phases	253
	7.2	Aims and Objectives and Assessment Objectives	
	7.3	Thinking in the Subject	
	7.4	Mark Schemes and Assessment	
	7.5	Limitations and Recommendations for Further Research	272
8.	Bibli	iography	276
9.	Арре	endices	

Table of Tables

Table 2.1 Piaget's Stages of Development 29
Table 4.1 Breakdown of Category Instances in Relation to Total Instances164
Table 4.2 Script Analysis: Categories as a Proportion of Total Instances.
Table 5.1 Descriptive Statistics for Bloom Mark Scheme, Question 1
Table 5.2 Descriptive Statistics for SOLO Mark Scheme, Question 1
Table 5.3 Comparing Ranking of Student Answers to Question 1 using theBloom Taxonomy and the SOLO Taxonomy.183
Table 5.4 Descriptive Statistics for Bloom Mark Scheme, Question 2
Table 5.5 Descriptive Statistics for SOLO Mark Scheme, Question 2
Table 5.6 Marker Comments for Student 5, Question 2, SOLO Mark Scheme 186
Table 5.7 Z- score Calculation for Question 1 Using Bloom and SOLO MarkSchemeScheme
Table 5.8 Standardised z-scores for Question 1
Table 5.9 Rank Order: Question 1. 192
Table 5.10 Standardised z-scores for Question 2. 193
Table 5.11 Rank Order: Question 2193
Table 5.12 Bloom Q1 Normality Test
Table 5.13 SOLO Q1 Normality Test 195
Table 5.14 Bloom Q2 Normality Test
Table 5.15 SOLO Q2 Normality Test 196
Table 5.16 Coefficient of Variation Question 1. 198

Table 5.17 Coefficient of Var	iation Question 2
Table 5.18 Coefficient of Var	iation by Student, Question 1202
Table 5.19 Coefficient of Var	iation by Student, Question 2203
Table 5.20 Sample Conversion	n and Contingency Table205
Table 5.21 Bloom Mark Schem	ne Mark Conversion206
Table 5.22 Example of a Cont	tingency Table206
Table 5.23 Frequency Scores	Contingency Table for Assessor 1206
Table 5.24 Assessor 1 Chi-squa	are Output208
Table 6.1 Concepts identified	by the sample218

Table of Figures

Figure 3.1 Raw marks for students in the sample	146
Figure 3.2 Coder Analysis	147
Figure 5.1 Coefficient of Variation for Question 1	199
Figure 5.2 Coefficient of Variation for Question 2	201
Figure 5.3 Coefficient of Variation by Student, Question 1	202
Figure 5.4 Coefficient of Variation by Student, Question 2	204

1. Introduction.

1.1 Background and context of the research.

The relationship between teacher and student is a classic example of asymmetric information; the student is often assumed to know very little whilst the teacher is the 'expert'. The task of the teacher is to guide students on a path to becoming more like an expert. Teachers will have already made such a journey, and it is often very difficult to remember what our understanding was at the start of the journey.

There are other motivations for the teacher in addition to that of developing growing expertise in the subject. This may include helping the student to gain a qualification in the subject and prepare for the next phase of the learning journey. Teachers may perceive that they are judged, not on the extent to which they develop their students' expertise in the subject but on the grades that they secure, which is used as 'evidence' that the department or school is providing a high quality education.

For students beginning an examination course at age 14, the start of the journey can be daunting. The asymmetry between teacher and student can be wide; the challenge of leading students on the path to becoming an expert or achieving a desired grade can be considerable and require a particular philosophical approach and conception of learning.

As students' progress through the journey they experience problems in understanding new terms and concepts and grasping the tools, concepts, methods, theories and models that underpin how a discipline is taught and understood by those who have travelled the path. The measure of the extent to which learning progress is made is through an examination. This thesis focuses on this initial journey and issues surrounding the assessment of student understanding and skills through the GCSE examination taken in England and Wales at age 16.

All students make a journey through learning but not all become 'experts' and come to see the discipline they are studying in a transformed way. The reason why some students do come to think differently as a result of their journey and why some do not has led to the notion of threshold concepts (Meyer and Land, 2003), which has received increasing attention from researchers over the last ten years. Most of this attention has been focused on higher education in a similar way to the research which has been conducted on approaches to learning (Marton and Säljö, 1984; Entwistle, 1983; Ramsden, 1992; McCune, 1998; McCune and Entwistle, 2000; Entwistle, McCune and Hounsell, 2002).

Students who pursue a course of study in a discipline area such as economics and business at GCSE do have some similarities with higher education students embarking on a degree. Undergraduate students will have chosen to pursue a course of study in a particular discipline. Some will have background experience of the subject area they have chosen to study but the nature of the application process in England and Wales means that it may not be a requirement to have studied the subject chosen for their degree prior to entry.

In the vast majority of cases in England and Wales, students make a conscious choice to take a GCSE in economics and/or business in the same way that students at age 18 choose economics or business as a degree course. Economics and business are not part of the National Curriculum and as a result most will not have had any formal exposure to economics or business as academic disciplines.

At the end of a degree course, it might be expected that students will have developed some expertise in the discipline they have chosen and it is no less to be expected that students at the end of GCSE might have developed some initial expertise in the subject. In both cases, students will have been subject to an assessment designed to measure the extent of the progress made in their understanding and to have provide some evidence of learning. The nature of this evidence of learning will be explored further in this thesis but at GCSE is based around assessment objectives defined by government and regulatory bodies and interpreted by awarding bodies who administer the examination and qualification. The difference in evidence of learning at GCSE and degree level will be in the extent of the detail and skills expected.

It is to be expected that some transformation of understanding will have taken place in both sets of students. They will have experienced new knowledge and developed new skills not possessed at the start of the course of study. In both groups, some students will have made considerable strides in developing expertise and will have 'got it'. Equally, some students in both cohorts will have struggled to come to terms with the tools, concepts, terms, models and methods of the discipline. Differences between student outcomes ought to be reflected in the grades which the respective assessments generate. Students who gain high grades at GCSE and who gain 2:1 or 1st class honours at degree level can be assumed to have developed a level of understanding and expertise that is qualitatively and quantitatively different to those who have achieved lower grades.

Research into higher education has revealed a number of issues which may cast doubt on this assumption. Such research has identified surface, strategic and deep approaches to learning (Entwistle, 1983; Ramsden, 1992; McCune, 1998; McCune and Entwistle, 2000; Entwistle, McCune and Hounsell, 2002). Research on lecturers in higher education has been equally extensive and has revealed different conceptions of learning (Ramsden, 1992; Prosser & Trigwell, 1999; Laurillard, 2002; Biggs, 2003). There has been doubt cast on the validity and reliability of terminal assessments, such as problems and essays, used in higher education (Heywood, 2000; Moss, Girard and Haniford, 2006; Rust, 2006; Sambell, McDowell and Brown, 1997). The reason why some students appear to 'get it' whilst others struggle to cope, evidenced by both formative and summative assessment, has led to the notion of threshold concepts (Meyer and Land, 2003) and troublesome knowledge (Perkins, 1999) as one possible explanation.

1.2 The nature of the research problem

My background has been as a teacher of economics and business and as an examiner at a senior level in the national examination system in England and Page | 13

Wales. I was aware of the research into the issues outlined above at higher education. My observations in the students I was teaching, in the conversations I had with colleagues and in the exam papers I was marking raised issues that were very similar but at a different level to those in higher education. There seemed to be little research into these issues below higher education. This provided the motivation for the research.

Interaction with a large number of teachers through training I conducted for the awarding body I was working for, highlighted an increasing focus on 'getting grades'. Targets set by government on levels of achievement for GCSE pass rates are used for promoting the school and the quality of education provided. There seemed to be an implied and assumed correlation that an increase in the proportion of A* to C grades at GCSE was a signal of improved quality in teaching and learning in the school. My observations of GCSE examination scripts suggested that the pressure teachers were feeling was translating itself into a focus more on 'teaching to the test'. The amount of information and feedback given to schools about the examination, the performance of students, mark schemes, access to past papers and examiners' reports seemed to help fuel this focus on grades and help teachers teach to the test more effectively.

Reports from the Office for Standards in Education (Ofsted), the inspection service, suggested that teaching in economics and business could be 'uninspiring', (Ofsted, 2008). The UK qualifications regulator, Ofqual, promotes 'thinking in the subject' in its subject criteria (Ofqual, 2011a). My experiences and the statements from Ofsted and Ofqual seemed to be at odds. To what extent was my observation of patterns in exam answers typical? Were teachers more focused on grades than on student learning and understanding and where does the idea of thinking in the subject fit in with GCSE assessment?

This thesis links research in higher education with my observations at GCSE level to identify possible issues that could lead to a reconceptualisation of the way assessments in economics and business at this, and other levels, are put together. It is postulated that a reconceptualisation could encourage a change in teaching and learning approaches and help to encourage the idea of helping students on the learning journey to developing expertise in their subject; to begin to think in the subject.

The notion of threshold concepts and troublesome knowledge offers ideas which explain why some students, even at GCSE level, seem to 'get it' (to demonstrate understanding, skills and thinking in the subject) whilst others continue to experience considerable difficulties. Economics and business are subjects which have a considerable number of abstract concepts and use language that is sometimes at odds with everyday use. In addition, the methodology of the disciplines uses models as a means of explaining complex issues and which, of necessity, make some assumptions which students often find difficult to accept.

Evidence of learning at GCSE is measured by the acquisition of a grade in a summative examination. The assessment has some influence on the teaching and learning programme which teachers design. In particular, assessment objectives are a crucial factor in determining what the student is expected to achieve and the skills which students are expected to demonstrate. The assessment objectives are detailed in the qualification specifications, published by the different awarding organisations. These provide details on the aims and objectives of the qualification, the content to be covered and the format of the examination paper/s. Teachers at GCSE will be conscious of the specification for the awarding organisation they have chosen and work closely to its guidelines in preparing their teaching and learning programmes. The assessment will, therefore, have a part to play in the way in which teachers approach their planning for the subject and may push a focus on particular aspects which enable students to gain the grades they need rather than a broader goal of beginning the process of thinking in the subject.

Observations made in my examining role suggested that understanding of the assessment objectives was not as consistent as might be assumed. One of the key aspects of setting and maintaining standards in examining is the

assumption that examiners can come to have a common and collective understanding of the standards set by senior examiners and use this understanding to mark exam papers consistently, both within exam series and across time. This assumption allows awarding organisations to make decisions on setting grade boundaries for examinations and awarding students with grades based partly on the judgement of the senior examiners but also on the statistics which are derived from the current and previous cohorts. If this assumption can be questioned then a foundation of the examining system could be in doubt.

If it is possible to question the assumption that examiners can be standardised even when they receive training, then it might also be reasonable to assume that the level of consistent understanding of the assessment objectives amongst teachers in general could be questioned. If teachers have different understanding of the assessment objectives this is going to have an effect on the way that they prepare students for the examination and the assessment.

1.3 Research Questions

This background led to a series of research questions being developed. These are:

- 1. To what extent do student answers in examinations show evidence of beginning to think in the subject?
- 2. Do teachers have a common understanding of mark schemes developed for assessments?
- 3. Do teachers have conceptions of learning which reflect deep or surface approaches to learning?
- 4. Would changing the assessment objectives in specifications lead to a different focus for teachers and encourage deeper approaches to learning and thinking in the subject?

1.4 Position of the Researcher

This thesis will be based on a mixed methods approach but with a greater

emphasis on qualitative which forms the basis of two of the three phases of the research. The design of phases I and III required the use of content analysis and online questionnaires and so qualitative methods were appropriate. Phase II required some statistical comparison of numerical data and so was quantitative in nature. The combination of qualitative and quantitative methods reflected the research questions and together allowed me to better address the overall research problem (Cresswell, 2015).

Positionality can be defined as the social and political landscape occupied by the researcher (Strathclyde, 2014). If another individual pursued exactly the same research as I have done, there is a probability that the interpretation would result in very different conclusions and recommendations. Both outcomes would be a 'truth' but truths generated by different researchers who bring diverse perspectives and experiences to bear on the data collected. The individual researcher is the ultimate arbiter of what they choose to include in their report and, crucially, what to leave out and in doing so they hold a particular position in relation to the study (Wood Wallace, 2012). To ensure that the conclusions and recommendations of this research are credible, accepted methods of analysis are adopted which mean the research is trustworthy.

It is inevitable that my experiences and observations will have some impact on the choice of methodology used, the way in which the data is analysed and the conclusions and recommendations generated. Stenhouse (1975), refers to this as the impulse behind all research. Bias, defined as being a particular leaning towards an interest in a topic or area, can be seen as an integrative feature of any research. Jones (2001), in discussing the debate on objectivity in research, notes that it is not a case of whether total objectivity can ever be reached but more that qualitative research can be meaningful if the purpose and the assumptions made by the researcher are clear.

I arrived at this research through an interest in individual differences which had been generated through observation of the different ways in which students demonstrated learning both in the classroom and through reading examination scripts over a twenty year period. These experiences led me to question the basis on which students were being judged and whether there was a different way in which learning could be measured which more accurately reflected the extent to which students understand and think in a subject. This would mean challenging the basis on which national examination assessments are based and specifically the focus on assessment objectives based on Bloom's Taxonomy (Bloom, et al 1956).

It is not surprising, therefore, that in conducting this research I might have a greater degree of sympathy for certain views and interpretations expressed by some respondents in Phase II and III of the research and in the way in which I coded and interpreted the content analysis in Phase I. An important element of the qualitative research I have undertaken is that where I see a respondent's comment or perspective which, on first reading, is diametrically opposed to mine, I do not simply dismiss it but look to consider the possible context in which the respondent is making the comment.

Educational research is focused on students and teachers and the perceptions of both will vary dependent on the context in which they are operating. Some respondents to Phase III noted that they teach in the private sector and the pressures and expectations which are associated with that sector may be different to those individuals working in the public sector. Both are valid perspectives but different. Each individual will have a set of 'truths' and the researcher is no different.

Equally, it is important to be aware of the potential for confirmation bias. Ross and Anderson (1982), define confirmation bias as the human propensity to process evidence in the light of existing held beliefs and the subsequent tendency to provide more support for those beliefs than the evidence might suggest. Kahneman (2011), charts the development of this idea from the seventeenth century philosopher, Baruch Spinoza. Kahneman (2011), expresses confirmation bias in the context of 'System 1' and 'System 2'. System 1 is the fast, intuitive and emotional response to external stimulus and is responsible for the initial impulse to believe and System 2 is the slower, more deliberative and more logical processing of that same external stimulus responsible for the questioning, doubting and disbelieving. As a researcher, I have to be aware of the potential dominance of System 1, to ensure that I engage System 2 and to explicitly state the assumptions and position from which I am analysing data.

My particular perspective on the teaching and learning of economics and business is one which is based not on solely preparing students for an examination but as the teacher as participant in the student's learning journey. I see it as part of my role to engender an enthusiasm and interest in the subject so that students leave the classroom and transfer the knowledge, understanding and skills they have been exposed to in their interactions with the outside world, their world - to 'think in the subject'. My perspective on teaching and learning and the assessment of that learning (my bias) is on encouraging a deep approach to learning. I perceive a link between deep approaches to learning and threshold concepts and troublesome knowledge and so my research focus includes these and seeks to interpret them at GCSE level.

In considering positionality, therefore, it is important to make the reader aware of the location of this researcher in the field and to reflect on that position in the discussion. As Griffiths, (1998, p.133) notes: "Bias comes not from having ethical and political positions - this is inevitable - but from not acknowledging them. Not only does such acknowledgment help to unmask any bias that is implicit in those views, but it helps to provide a way of responding critically and sensitively to the research."

1.5 Limitations of the Research

The discussion and recommendations section of this thesis contains suggestions for the reform of assessment objectives which, in turn, could be used as the basis for curriculum and assessment development and influence the teaching and learning approaches of teachers and students. Such an outcome represents a lofty ambition and the recommendations will be constrained by the limitations of this research. Research is limited by the capacity of the researcher and the extent to which I have been able to access an appropriate range and depth of material in the field work. As a part-time researcher with limited funds and a full-time job, the opportunities to access wider samples of data in all three phases have been a major challenge. The small sample sizes used in each phase of the research, the difficulties of analysing the data generated in Phase II and the extent to which any results can be generalised are all important limitations of this research. The examination papers used in Phase I were from one awarding organisation and so the extent to which the interpretations of the results of this sample can be extended to other awarding organisations and other qualifications apart from GCSE is open to debate. In Phases II and III, the research was limited by a trade-off; the desire to get some useful data to analyse against the knowledge that if more detailed and time-consuming contributions from respondents was required it was likely that the response rate would have been even lower than it was. Response rates to both phases were disappointing and the resulting samples subject to a greater degree of skewness in comparison to a more formal random sample selection process. Indeed, it can be argued that the respondents who did bother to reply might necessarily be those who do have an interest and enthusiasm for teaching and learning which will introduce some bias into the research. Despite these reservations, the findings of the research do lend a perspective to the debate on the nature and aims of assessment and provide some useful and original points for future debate and further research.

1.6 Outline of Chapters

Chapter 1 provides some background to the research, the motivations for the research and the main research questions which the thesis is attempting to address. There is a discussion of positionality which seeks to outline some of the considerations that need to be taken into account when reviewing qualitative data and the place of me, as the researcher, in the field.

The literature review will focus on five main areas:

1. Theories of learning

- 2. The purpose and nature of assessment and assessment objectives
- 3. Approaches to learning
- 4. The nature of concepts and thinking in the subject
- 5. Threshold concepts and troublesome knowledge.

Chapter 3 outlines the methodology used in this thesis and discusses the challenges, benefits and limitations of mixed methods research and the issues that arose in the collection of data in three phases of field work conducted.

Chapter 4 deals with Phase I of the research which comprised content analysis on a sample of examination scripts. Chapter 5 covers Phase II of the research which involved a marking exercise with a sample of economics and business GCSE teachers to get some insight into the extent to which there was common understanding of assessment objectives and mark schemes. The data collected was analysed using a mixture of quantitative and qualitative methods. Chapter 6 covers the final phase of the field work which involved an online questionnaire designed to elicit information on their conceptions of learning. The answers to the questions were subject to text analysis. Chapter 7 pulls together the three phases of the research and discusses the results, the inter-relationships between the three phases, the extent to which the results address the research questions posed, what new contribution to knowledge the results suggest and what recommendations can be made for the reconceptualisation of assessment objectives and the nature of assessment at GCSE to better encourage the development of students on their learning journey to think in the subject. These recommendations have some implication for further research into the association between the way an assessment is designed and the approaches to learning adopted by students and conceptions of learning of teachers.

1.7 Summary of Chapter

This chapter has outlined the motivation for the thesis and how my background and experience as a teacher and examiner led to a series of questions which the thesis seeks to address. It considers some of the methodological imperatives underpinning the research which will be more fully explored in Chapter 3 and presents the overall structure of the thesis.

In the next chapter, the literature review will consider key contributions in a variety of distinct but related fields of research, to provide a survey of the current thinking in the different research areas I am exploring. This provides a benchmark for the contribution that this thesis will make to the field and in particular the application of the research fields to GCSE teaching and learning and assessment.

2. Literature Review

2.1 Theories of Learning

One thing common to every human being and to a considerable extent to any member of the animal kingdom, is the capacity to learn. Given this fundamental commonality it is perhaps surprising that there is a diverse range of theories of learning. In reviewing the literature on theories of learning, it is pertinent to establish a definition of learning that will underpin the discussion throughout this thesis. Learning can mean different things in different circumstances and it is likely that there is not one theory that adequately covers all instances of human learning.

Learning can be described in terms of the everyday necessities that help us to function effectively. This starts at birth, or even before, and involves such momentous milestones as learning to crawl, walk, talk, appreciate danger, how to eat, wash ourselves and so on. From the age of 3 or 4, we come into contact with a different sort of learning, a learning that is, to a large extent, imposed on us for our own good. We learn how to read, write, understand sentence structure, the basics of number and later, basic principles in science, geography and technical subjects. Curry (1983) refers to this learning as intended learning. Whilst Curry (1983) does not give a specific definition of what is meant by intended learning it is taken to mean the process whereby a body of knowledge is made available to an individual with the intention of that individual demonstrating mastery and understanding of that knowledge to different degrees.

Honey and Mumford (1986) describe learning to have occurred when an individual knows something he (*sic*) did not know earlier and can show it; s/he is able to do something s/he was not able to do before. Kolb (1984) sees learning as a dialectic tension and conflict between the immediate concrete examples held by one group, which could be the student, and the analytic detachment of another, which could be the teacher. The bringing together of these two experiences challenges the perspectives and opinions of both

groups to create vitality and stimulation. These definitions of learning are representative of different aspects of learning which will be covered.

This thesis is based upon the process of intended learning in schools at GCSE. Students come to a body of knowledge and skills in a subject that will lead to a change in the reality that existed before. This will lead to the student being able to make connections between unrelated pieces of information, deal with change in a confident and competent manner and will allow the student to be in a position where they can make sense of the world around them, a world that has some meaning to them as individuals.

This discussion about what learning means is important because what learning actually means in the context of academic study can be taken for granted. We do not always need to learn something in depth whereas on other occasions it may be essential to enable the student to move forward in the subject. (Bowring-Carr and West-Burnham, (1997), p 107) highlight these differences thus: "The taking in and regurgitating of facts, is instantly forgettable; it has no impact on self, and therefore no impact on behaviour. Children can get ten out of ten on any number of tests on 'What should you eat to be healthy?', and then immediately go out and gorge cheerfully on the junkiest of foods."

The following represent aspects of a definition of learning which will form the basis of this thesis:

- Learning is about acquiring new knowledge AND skills 'knowledge is the know why, skill the know how' (Boyett and Boyett, (1998), p85)
- Learning is about changing behaviour as a result of the learning; this new behaviour is underpinned by a different set of assumptions about 'reality'.
- Deep learning must involve, and have meaning to, the individual
- Learning means making connections between unrelated information and making connections
- Learning is about being adept at dealing with change

- Learning is about being creative and finding solutions to problems
- Learning is about being independent but also about being able to work well in communities.

For the purposes of this thesis learning will be defined as a change in thinking as a result of the acquisition of new knowledge, skills and understanding. Knowledge in this context, is defined as a process of how we come to know things (Hilgard and Bower, 1975).

2.1.1 Historical Perspectives on Learning

How people learn and what methods lead to the most effective learning has been the subject of debate for thousands of years. The epistemological origins arise out of discussion about what sort of learning is considered appropriate and what purpose learning serves. For Plato, learning and knowledge was the process of retrieving what was already known and acquired in a previous life (Spender, 1996) but for Aristotle learning and knowledge came from empiricism, searching for truth and knowledge in the world outside the individual (Darling-Hammond et al, 2001) and through doing (Aristotle, 2009). Philosophical debate about learning and knowledge continued through Hobbes (1588 - 1679), Descartes (1596 - 1650), Locke (1632 - 1704) and Rousseau (1712 - 1778), (Hilgard and Bower, 1975) but the focus on the scientific study of learning blossomed with developments in psychology.

2.1.2 Behavioural Theories of Learning

The focus on changed behaviour as a fundamental part of learning dates back to the early development of psychology as a discipline where researchers questioned why humans behaved as they do. Behaviourism developed out of a move to make research in psychology more akin to that in the sciences where methods of inquiry were capable of being replicated and subject to objectivity (Nolen-Hoeksema et al, 2009). The response to external stimuli was more capable of being controlled and measured than introspective mental causation, which had characterised much research in early psychology (Nolen-Hoeksema et al, 2009). Psychology was seen as a science of behaviour (Skinner, 1950) and behaviour, in turn, seen as being strongly influenced by stimulus and response. Thorndike (1923), saw the existence in 'man' of 'original tendencies' and 'environment' as factors influencing behaviour and that this was 'elementary psychology'. Humans responded to situations which made connections in the brain and led to changes in behaviour. Thorndike (1923, p70) refers to learning as "permanent modifiability", and the aim of education as perpetuating [connections] to generate the desired behaviour whilst eliminating, modifying or redirecting other connections. The strength of the connection between the situation or stimulus and the response was determined by a series of 'Laws of Learning' (Thorndike, 1923). The Law of Use stated that the more use between the situation and response, the stronger the connection and the greater the learning. The 'more attentive', vigorous and duration of the situation and the response, the stronger the connection. The Law of Effect noted that modifiable connections followed by a satisfying state of affairs strengthened the connection. Thorndike (1923) describes experiments with chicks and kittens where the response of the animals to different situations was observed. He noted that repetition facilitated the forming of connections which influenced behaviour.

The association of learning with stimulus and response was further developed by B.F. Skinner. Skinner developed the theory of instrumental or operant conditioning (Skinner, 1948), the heart of which involved observations of the relationship between the response to external stimuli and the results of the response. Skinner's work with rats and pigeons suggested that appropriate external stimuli, such as the provision of food, reinforce particular behaviours such as the pressing of a lever which releases food. However, when the external stimulus is removed, the behaviour changes and the pressing of the lever is reduced. Skinner concluded that positive reinforcement increases the likelihood of the desired behaviour and that negative reinforcement can reduce a behaviour.

The work of behavioural psychologists implied that learning could be seen as

being the generation of appropriate and/or desired responses to stimuli and that desired behaviour could be reinforced through repeated positive experiences. The consequence of the behaviour occurs *after* the behaviour and is less likely in the absence of the behaviour. The relationship between the behaviour and the consequence changes the future probability of the behaviour (Purdy et al, 2001).

In the context of teaching and learning in school, if a student answers a question, submits a piece of work or takes a test and gets an affirmative response from the teacher, these positive reinforcements will encourage the student to study and increase the propensity to learn. If, on the other hand, the student is ridiculed for asking 'stupid' questions, or receives negative feedback on work submitted, the student will find study and learning unpleasant and be less inclined to learn. Such a view of learning has been referred to as 'skill and drill' (Graduate Student Instructor Teaching and Resource Center, 2014).

Behavioural or associative theories of learning have some value in the development of certain types of learning where mechanistic tasks are required or where students need to acquire knowledge and skills which can be honed through repetition and practice. Instrumental or operant conditioning, therefore, works best when associated with voluntary behaviours (Cacioppo and Freberg, 2013). Behavioural theories have limitations when more complex cognitive processing is the desired outcome of learning. Such limitations led to the development of what can broadly be termed cognitive theories of learning.

2.1.3 Cognitive Theories of Learning

The challenge to behavioural theories was led by the so-called gestalt theorists of which Köhler and Koffka were two leading researchers. Köhler carried out experiments on apes between 1913 and 1917 and his conclusions were counter to those of Thorndike. Köhler (1925), suggested that insight was a factor in the way some animals approached problems and that some element of cognitive process was being applied. Köhler's apes could see the

Page | 27

value of using tools to acquire food and used perceptual insight to learn something Hilgard and Bower, (1975) refer to as a *eureka* moment.

Gestalt psychology focused primarily on understanding perception applied to learning. Memory consists of the remnants of perceptual events so when faced with problems, memory is utilised to 'see' and to select the appropriate approach to solving the problem - a definition of insight (Hilgard and Bower, 1975). Such higher cognitive processes meant that experience of the world was in the form of creating meaningful patterns or as an organised whole. Knowledge is organised to solve a problem and is grouped into elements according to principles called 'The Laws of Organisation' (Cacioppo and Freberg, 2013). These principles are proximity, where elements are grouped in patterns according to nearness to each other; similarity, where items similar in some respect are grouped together, although differences can be discerned; closure relates to the grouping together of items to complete some entity and simplicity where stronger or adequate patterns dominate weaker patterns in perception. Where symmetry, regularity and smoothness are dominant, we organise simple figures according to these dominant patterns (Cacioppo and Freberg, 2013).

The role of cognition in learning was further developed through the work of Piaget and Vygotsky. Jean Piaget (1896 - 1980), based his work on observations of children and suggested that learning was a developmental process. At certain stages of development, children are able to handle different physical and mental tasks but some cognitive tasks prove to be beyond them. Piaget saw human learning as an adaptive feature that enables humans to deal effectively with their environment, where conceptions of reality are shaped by our continuous interaction with the environment (Cacioppo and Freberg, 2013). Piaget saw knowledge not as a body of information an individual acquires or a state of possessing information but conceived of knowledge as a process which required mental and physical action (Murray Thomas, 2000). Knowledge growth is seen as a sequential process of logically embedded structures (schemata) (Cacioppo and Freberg, 2013). Table 2.1 summarises the various stages of development from birth to maturation.

Age	Development Stage	Abilities*
0 - 2	Sensorimotor	Child's world is of the 'here and now' - objects exist only when the child senses them. Child understands the environment through sensations and actions towards it. As the child develops, they begin to symbolize and communicate.
2 - 7	Pre-operational	Objects come to be represented internally but similar objects still seen as identical (for example, the child may see several Santa Clauses in one day and can recognise each as 'Santa' but does not distinguish that they are all different Santas. As the child develops, thinking becomes more logical but subject to the child's own perspective (what Piaget called egocentrism).
7-14	Concrete operations	Thinking becomes more 'rule-regulated' – as the child interacts with its environment it realises that logic governs actions and relationships.
14+	Formal operations	Ability to think about abstract relationships, understand methodology, formulate hypotheses and think about abstractions.

Table 2.1	Piaget's	Stages of	Develo	nment
	i layet s	Judges Of	DEVEIU	pinent

*from Lefrançois (2012)

Piaget's developmental stages have some relevance to the learning of students in economics and business at the age level covered by GCSE qualifications. Economics and business content includes a number of abstract concepts and the ability of learners to grasp these concepts when their cognitive capacities have not matured sufficiently, or at the very least are maturing at different rates, presents additional challenges for teaching and assessment.

One particular relevant implication from Piaget's work to the learning of new discipline areas such as economics and business at GCSE is that children experience their environment using existing mental maps. If the child's experience draws on, and fits into the child's cognitive structure, it can result in the maintenance of a mental equilibrium which Piaget referred to as *assimilation*. When children are exposed to new and different experiences this challenges the mental equilibrium and cognitive structures have to change to accommodate these new conditions. As a result the child builds more new and more adequate cognitive structures (Cacioppo and Freberg, 2013). The relevance of this is exemplified in Phase III of the research where respondents focused on the challenges that students face in coming to terms with new information and concepts in economics and business which they have not encountered before because few students get the opportunity to study these disciplines prior to age 14.

Learning does not occur in isolation and the social-cultural context of learning was emphasised by Vygotsky. Vygotsky (1978), noted that classical psychological literature assumed development is a prerequisite to learning and the lack of mental maturation meant that any amount of instruction would be of little benefit to learning. Teaching effort is focussed "on finding the lower threshold of learning ability, the age at which a particular kind of learning first becomes possible" (Vygotsky, 1978, p30). People develop in the context of a culture and human cognitive development is affected by the culture in which people are brought up in. Culture teaches children what and how to think, and they will mould their behaviour on the observed behaviour of others, most notably parents. Learning is, therefore, dependent on social interaction (Cacioppo and Freberg, 2013).

Vygotsky (1978) proposed that learning could influence the maturation

process. By the time children begin school they will have already experienced diverse learning situations which influence the abilities children possess and the rate at which new learning can be absorbed. Vygotsky distinguished between pre-school and school learning as non-systematic and systematic. 'Systematicness' was not the only issue; learning needs to be matched with the child's developmental level, both the actual development level (the mental functions established as a result of completed developmental cycles), and what he termed the 'zone of proximal development', (Vygotsky, 1978). The zone of proximal development is "the distance between the actual development level, as determined by independent problem solving, and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p33). The concept of the zone of proximal development emphasised the use of written and spoken communications and cooperative learning and led to the development of the idea of scaffolding, central to constructivist learning theories, (Darling-Hammond et al, 2001). Vygotsky's work added a new dimension to the work of Piaget whose stages of development theories have been referred to as 'crude' and introduced the affective domain as a central element in the cognitive domain (Jarvis et al, 2003).

The social dimension of learning was further developed by Bandura (1977), which he termed social learning theory. Bandura saw learning as a social process and one of reciprocal determination. He suggested that people are not simply reactors to external influences – the individual and environmental influences are interdependent. "People's expectations influence how they behave, and the outcomes of their behaviour change their expectations" (Bandura, 1977, p195). It follows that people can influence their own destiny as well as the limits of self-direction. Human functioning does not imply that they have no self-direction but equally they are not totally free to become whatever they choose. People and their environments are reciprocal determinants of each other (Bandura, 1977).

Bandura (1977), also identified that anything can be learned by direct

experience but can also be learned through observation. Teachers can act as models which can be effective if they command respect and have competence, status or power. Teachers can carefully plan what is presented and as a result do more than merely teach routine information. Model behaviour becomes the standard for student self-evaluation. (Olsen and Hergenhahn (2009).

2.1.3.1 Constructivism

This thesis assumes a constructivist theory of learning which can be both individual and social. Constructivist theories of learning are a subset of cognitive theories. Constructivists see knowledge as a constructed activity. The learner experiences the environment in many different ways and reflects on these experiences to construct a perception of reality. Constructivists see learning as not simply the transmission of information from one individual or resource, such as a textbook, to another but a process where individuals build their own meaning and understanding of information. Learning is, therefore, a search for meaning and of necessity requires personal interpretation (Martin and Loomis, 2014).

Given that constructivist theories focus on the processes whereby learners build their own mental structures when interacting with the environment, the pedagogical focus is task-based and favours hands-on, self-directed activities oriented towards design and discovery (Lefrançois, 2012). Constructivism implies a structuring of learning environments, such as simulated worlds, so as to inform the construction of certain conceptual structures through engagement in self-directed tasks. (Illeris, 2009)

Constructivists view reality and meaning as personally rather than universally defined. The assumptions of constructivism emphasise content, context and understanding, the individual negotiation of meaning and the construction of knowledge. Constructivists favour rich, authentic learning environments rather than isolated decontextualised knowledge and skill transmission. The emphasis is on student-centred, goal directed inquiry rather than externally directed instruction and supports personal perspectives over canonical

perspectives (Jonassen and Land, 2000).

Constructivist knowledge implies that the knower is an intimate part of the known and that questions and answers vary depending on the historical and cultural context and the inquirer's frame of reference. Posing questions and problems become the key methods of enquiry (Jarvis et al, 2003); constructivists search for truth beyond and across systems, a characteristic of deep learning.

There are a number of branches of cognitive and constructivist thinking with each placing a greater emphasis on a factor which influences learning such as society, culture and environment. Bruner (1966), emphasised the role of cultural and narrative experiences in learning. Emphasis in education, therefore, was to be focused on skills and symbolic operations, especially in regard to technology that benefits and enhances cultural expression. This emphasis implies that most learning in most settings is a sharing of culture. (Bigge and Shermis, 1999). Bruner's concern is with the means by which people select, retain and transform information, to generalised insights and understandings, (Bigge and Shermis, 1999). Bruner, (1966) refers to folk psychology, the knowledge of a system whereby people organise experiences in, knowledge of and interactions with their environment - the way people make meaning. Bruner, (1966) suggested learning involves three almost simultaneous processes, acquisition of new information or a refinement of existing knowledge, transformation of knowledge - where knowledge is manipulated to make it fit with new tasks or is converted to another form, and a check of the pertinence and adequacy of knowledge by evaluating whether the way we manipulate knowledge is adequate to the task in hand including judgements of the plausibility of the knowledge, (Bigge and Shermis, 1999).

Bruner's processes imply that learning should be structured, rich and with an emphasis on the conceptual framework - the bigger picture. Learning principles are then discovered by the learner themselves and the teacher's role is to develop the big picture, (Bigge and Shermis, 1999). Learning

Page | 33

becomes something where there is some benefit gained which Bruner (2006) likens to leaping over a barrier and on the other side of the barrier is thinking. Past learning has to be organised so that it is not bound to the particular context or situation in which it was originally learnt but can be extended to new contexts and situations (Bruner 2006).

The debate on the extent to which knowledge is absolute or personally and socially constructed has also been played out in mathematics. Ernest (1994), states that mathematics education encompasses a tradition which assumes mathematics knowledge pre-exists and can be discovered by the learner. This assumption derives from a positivistic epistemological paradigm that characterises mathematics, particularly at high levels of study, (Ernest, 1994). However, when students discover mathematical knowledge they construct different meanings which can be fallible. Constructivism allows for the idiosyncratic, the systematic error, misunderstanding, misconception and alternative conceptions that arise in the interpretation of mathematical knowledge is one of the ways in which the boundaries of knowledge are pushed by individuals thinking in different ways, making the leap across Bruner's barrier and arriving at new interpretations of knowledge which advance our understanding of the world.

The constructivist view of learning, and of the pedagogy and assessment that is associated with it, has some connections with threshold concepts and thinking in the subject, which will be explored later in this chapter. von Glasersfeld (1994) draws attention to the centrality of the use of symbols in mathematics and how those who become used to arithmetic have grown to see little difference between the symbols and the concepts to which they refer. They have moved through the portal and have a transformed understanding of these symbols and their learning is now difficult to unlearn. von Glasersfeld (1994) notes that symbols have to be deliberately associated, through a mental process, with a conceptual meaning and there is no automatic concept generation by learners on first experiencing these symbols. Because these mental processes are individual, von Glasersfeld (1994) argues that they cannot be 'witnessed' by anyone else. He suggests that the consequences of this analysis is that teachers have to pay attention to mental operations of students and that teaching has to be concerned with understanding rather than rote-learning.

The next section of the literature review will look at the purpose and nature of assessment and assessment objectives. This forms the basis for understanding the learning that is being sought by the government, sets the criteria for the awarding organisations in developing the specifications which detail the content and learning to be assessed, and by teachers who have to interpret the criteria and assessment objectives and choose how to convey this content and understanding against the assessment framework in the classroom.

2.2 The Purpose and Nature of Assessment and Assessment Objectives

Theories of learning provide a means of predicting how people learn. Based on such theories, ways of measuring the extent of the learning following a particular form and style of instruction become possible. Almost all education, particularly the education of young people in their teenage years, focuses on finding ways of measuring learning outcomes. These measures are used for a variety of reasons. This section of the literature review will look at some of the features and characteristics of assessment and explore two taxonomies which each provide a different focus on the nature of the learning outcome. This is important because the nature of assessment which students face also influences conceptions of teachers and students and the nature of instruction. It follows that this can also influence the approach to learning of both students and teachers and whether the predominant approach is of a surface or strategic type rather than deep learning.

2.2.1 Definitions of Assessment

To assess learning there is an assumption that there is something that can be

measured and in such a way as to distinguish how much is possessed by different individuals (Killen, 2005). A clear definition of learning, in terms of what is being measured, how it will be measured and the subsequent recording and usage by those involved in the learning process, is necessary if the measure is to have credibility.

Assessment is the gathering of evidence of the desired measure and the subsequent interpretation about that evidence and its use to make judgements about students' achievements following a course of instruction, (Harlen, 2007). In the context of this thesis, the course of instruction is a two-year programme of study in economics and business. The assessment comes at the end of the course of study and is in the form of an examination. Student answers to questions in the examination provide the evidence of learning which is interpreted to provide a measure of the extent of the learning. This evidence has to be evaluated against a standard which is generally accepted as being appropriate for the level of learning being considered (Scottish Qualifications Authority, 2014). The standards set for students studying for a GCSE in economics and business takes into account the expected learning for students over a two-year course of study, typically between the ages of 14 and 16.

It has been noted that theories of learning identify stages of development of humans so it follows that there are going to be accepted limitations to the extent of the learning which is possible for the average 16-year old at the end of two years' study of economics and business. Students might not be measured on whether they demonstrate the qualities and characteristics of an expert in the subject, as might be the expectation at the end of an undergraduate programme of study, but how far they have begun the journey of developing an understanding and empathy for the subject they have studied.

2.2.2 Reasons for Assessment

Biggs and Tang (2007) identify a number of reasons for assessing student learning in the context of higher education. These reasons are given as:
- Selection
- Control or motivation
- Satisfying public expectations on standards and accountability
- For formative feedback
- Summative grading

All of above can be applied to the context of GCSE. The number, type and level of achievement at GCSE, expressed as the grades awarded for each examination sat, can be used to determine whether a student can access post-16 education at particular levels, either at advanced level (A-level) or through vocational education or, if they choose to enter the world of work, what sort of job can be accessed. In schools, assessment at GCSE is used for motivation and control to give students a focus for their studies as part of their education, opening up opportunities for them in later life. Having 'good' GCSEs means that opportunities to access higher education becomes more possible and with it the promise of higher life-time earnings.

The assessment system in England and Wales is also used for satisfying public expectations on standards and accountability representing the summative grading system on which schools, teachers and students are judged. GCSE results are used to create league tables and measure the performance of schools providing information to parents to make choices about which schools are 'good'. Teachers place importance on GCSE results as a measure of their teaching abilities and performance. The percentage of GCSEs acquired by students, including those in Mathematics and English, is the basis for the comparison of schools and changes in these percentages over time are used as a measure of the effectiveness of the school in 'delivering education'. If performance at GCSE level falls then schools and teachers can be held accountable and in some cases remedial action is put in place on recommendation from the schools inspectorate.

Within schools, assessment in its broader sense (i.e. not as a national examination) is used for both formative and summative purposes. Summative

assessment is used to form an overall judgement of achievement and report this achievement at a particular point in time (Black, 1999, Harlen, 2007). GCSE examinations are one form of summative assessment but over the period of the course of study, teachers will use summative assessment internally as a means of monitoring and reflecting standards of achievement at a particular point in time. For example, teachers will establish a series of 'end of topic' tests and mock examinations which may be used as part of the reporting process for students and parents and in making decisions about setting and alerting students about the size of the task facing them as they prepare for their final exams. Summative assessment is known as 'assessment *of* learning'.

Formative assessment is referred to as 'assessment *for* learning' (Black and Wiliam, 2003), and involves the short-term collection and use of evidence from teaching and learning programmes for the guidance of learning and informs decisions made about making progress in student learning (Black, 1999, Harlen, 2007). Formative assessment has become important in education in England and Wales and at GCSE. A recognition that students learn at different rates and in different ways has prompted a focus on the use of formative assessment in helping individualise learning and progression. Research by Black and Wiliam (1998), which reviewed a large number of studies on formative assessment, suggested that the effect on student learning was considerable to the extent that the gains could be a half to a full standard deviation.

The use of both summative and formative assessment influences attitudes and conceptions of both teachers and students. In some cases the divide between the two is blurred. For example, teachers will use tests as a means of establishing learning at a point in time (summative assessment) but use test results as a means of putting in place tactics and strategies to help students progress (formative assessment). In Phase III of this thesis, some teachers questioned outlined the regular use of tests for internal, school assessment. This blurring of formative and summative assessment and the methods chosen

to assess students can have considerable effects on the approach to learning by both teachers and students. Black (1999) notes that tests encourage rote and surface learning, that quantity is emphasised over quality, that marks and grades given are over-emphasised and advice and learning function underemphasised. Ramsden (1992, p18), states that "assessment defines the curriculum". This can frequently be an unacknowledged, invisible but powerful factor. Biggs and Tang (2007) point to the predilection of students to learn what they think they will be tested on, which they refer to as 'backwash', and as a result the assessment determines what and how students learn more than does the curriculum. Davies and Brant (2006) also refer to the idea of backwash as affecting the way in which students and teachers approach learning. If assessment focuses on individual parts of a syllabus then teachers will train students to display a more fragmented capability rather than holistic understanding (Davies and Brant, (2006). Biggs and Tang (2007), note that teaching to the test is the consequence of an exam dominated system which encourages surface learning and refer to this as negative backwash. For the teacher, summative learning may be seen as being at the end of the course of study but to the student it is at the beginning. Backwash can be also be positive if assessment is aligned to what students should be learning. This requires that learning objectives are clearly reflected in the assessment and that these objectives are linked to what the teacher intends that students learn. Jephcote and Abbot (2005) also note the importance of backwash. They suggest that teachers view the success of students (and by implication, their own success) in terms of grades. The content they teach may not stray far from the syllabus, even if the learning that might result would be beneficial, that teaching styles might crowd out the room for learners' inquisitiveness and that assessment strategies used in the classroom might only be those used by the awarding organisations that set the final summative assessment. This teacher-led approach, they argue, is reflective of behavioural psychology (Jephcote and Abbot, 2005). Dorman et al (2006), provide research which notes that teachers feel a need to sacrifice learning with understanding for the goal of drilling students in the things for

which they will be held accountable.

Black (1999), similarly argues that the link between assessment systems such as GCSEs, is a product of behavioural psychology. He argues that there is an assumption that learning can be broken up, taught and tested separately and that the test item is the stimulus to which the student is encouraged to respond appropriately, a point which is explored in Phases I and III of this thesis. Tests, Black (1999) notes, consist of short, atomised, out of context questions which encourage 'teaching to the test'. As a result, students are not encouraged to develop basic conceptual models or see meaning to what they are doing. 'Bits' become disconnected and students are unable to apply what has been memorised. There will be occasions when rote learning is necessary and important but this has to be balanced with the alignment of the goals of the final assessment (the GCSE examination in this case) and the broader goals of the qualification, which I argue should be based on thinking in the subject.

2.2.3 The Nature of Assessment

The measurement model of summative assessment such as GCSE examinations assumes, and is designed to measure, stable traits and abilities across a graduated scale to enable comparisons to be drawn between individuals. This implies that there has to be a means of referencing the outcomes and there are a number of ways that this is done.

2.2.3.1 Norm-referenced Assessment

Norm referencing involves comparing the performance of students against that of other students in the same cohort, age group, level of experience and/or on the same test, by placing students on a scale of achievement and comparing individuals with the 'norm' (Killen, 2005, Harlen, 2007, Wragg, 2001). Norm-referencing assumes that some students will find the assessment more challenging than others due to the natural results of motivation, ability, support, background and so on. Assuming a uniform marking system, students can be placed in a rank order from which a group norm is calculated, usually using the arithmetic mean. (Killen, 2005). Tests are constructed to spread Page | 40 students over a normal distribution. If the mean mark is (for example) 50%, most students will appear between 40% and 60% and much smaller percentages at the extremes of 1% - 20% and 80% - 100% (Wragg, 2001). Individual students can be compared against others in the group and their rank and position compared relative to the class average, (Killen, 2005).

One criticism of norm-referencing is that it can suffer from 'garbage-ingarbage-out' (GIGO). If marks are not gathered fairly and reliably or interpreted in valid ways then the marks can be meaningless. Rank ordering can also mask important aspects of what students have or have not learned. Wiliam (2000, p14) notes that "you can very easily rank order people without any idea of what you are putting them in a rank order *of*".

2.2.3.2 Criterion-referenced Assessment

Criterion-referenced assessment compares the performance of the individual against pre-determined criteria (Killen, 2005). The criteria could be listed in terms of what individuals should know or be able to do after a course of study. In some cases, in order to be deemed to have 'passed' the test, all the criteria have to be met in order to avoid individuals passing the assessment by performing at a high level for some of the criteria in order to compensate for poor performance on other criteria (Wragg, 2001), and in other cases, it might be possible to perform at a certain level across the criteria and gain a pass but higher levels of achievement could be secured if performance against the range of criteria is at greater levels.

There is an assumption that it is possible to describe the learning or the abilities that are to be achieved and that it is possible to have a cut-off point between what is an acceptable level of performance which constitutes a 'pass' and a level just below which represents a 'fail' (Killen, 2005). The criteria could be defined in terms of assessment objectives such as knowledge and understanding, application of knowledge, analysis and evaluation. Providing a clear definition of what each of these objectives mean and look like is difficult and open to interpretation. Students may also demonstrate a combination of skills in an answer in a GCSE examination, for example, at

different levels and there has to be some agreement on what these levels might look like, normally represented by a mark scheme. Even if the mark scheme is detailed it might not follow that all those who are marking the assessment will understand and interpret the mark scheme in the same way and thus share a common understanding of the criteria. The result may be that different assessors may make different judgements about the same performance, a problem which is highlighted in Phase II of this thesis.

The criteria also has to be relevant to a particular social context (Killen, 2005). At GCSE a criteria specifying the presentation of a 'complex and coherent argument' may require a different level of expected performance than that at A-level or from a first-year undergraduate. In other words, the criteria for a GCSE assessment has to be given in the context of what could reasonably be expected of the 'average' 16-year old student at the end of a two-year course of study.

2.2.3.3 Standards-referenced Assessment

Some of the difficulties with criterion-referenced assessment have led regulators to place an emphasis on standards. Ofqual has made it clear that assessing people's knowledge and skills in terms of standards is 'at the heart of what we do' (Ofqual, 2013). Standards referenced assessment compares learning to detailed descriptions of different levels of achievement which represent the standards (Killen, 2005). It is assumed that it is possible to define a range of levels of achievement of each learning outcome or groups of related learning outcomes. Assessment tasks are developed with the aim of enabling students to demonstrate what they have learned and compare students' achievements with a pre-defined range of possible levels of achievement. Students are thus judged by comparison with the detailed standards and placing students on a continuum of achievement (Killen, 2005).

Ofqual (2013) identify three types of standards, content standards, assessment standards and performance standards. Content standards relate to the syllabus content to be learned as part of a programme of study and how demanding this content is in relation to the depth of treatment and the level of skills involved. Assessment standards is the process of judging the extent to which students have met the demands of the assessment and what grade they can be awarded based on the evidence at hand, i.e. the student's answer. Performance standards at GCSE relate to the threshold or grade boundary which separates performance at one level from that of another.

There is some considerable overlap between criterion-referenced assessment and standards-referenced assessment at GCSE. Awarding organisations include the aims and objectives of the assessment in their specifications which tend to be generic. For example, Edexcel's Business Specification 2012, has as its aims to enable students to "actively engage in the study of business and economics to develop as effective and independent students and as critical and reflective thinkers with enquiring minds" and "use an enquiring, critical approach to distinguish facts and opinions, to build arguments and make informed judgements" (Edexcel, 2012, p11). It also identifies some key skills which the specification expects students to develop as a result of the course of study which include time management, organisation and planning skills, interpersonal skills such as group work and cognitive skills such as reflection and the review of own and others' performances (Edexcel, 2012). The assessment criteria (or assessment objectives) are split into three groups referred to as Assessment Objective 1, 2 and 3 (AO1, AO2 and AO3). The assessment objectives relate to the selection, recall and communication of knowledge and understanding, the application of skills, knowledge and understanding in different contexts and the analysis and evaluation of evidence to make reasoned judgements and present appropriate conclusions (Edexcel, 2012). The assessment objectives are weighted 35%:35%:30%.

The language used in the assessment objectives of all the awarding organisations offering GCSEs in economics and business subjects is nonspecific and open to interpretation. For example, what is meant by 'understanding', 'reasoned judgement' and 'appropriate conclusions' is not made clear. It is the job of the mark schemes to provide greater clarity.

Later in the specification, teachers are provided with a set of grade

descriptors (see Appendix I) which provide the expected standard of performance at three levels or grades, A, C and F. These grade descriptors are provided by Ofqual and apply to all awarding bodies. At grade awarding meetings, these descriptors provide a benchmark of the standards against which candidates' scripts are judged. A candidate might score 70% on the examination, for example, but whether this mark is worthy of a grade A has to be determined in part by the judgements of senior examiners, who are meant to view candidates' responses in comparison to the grade descriptors to decide whether the script/s meet the various skills and knowledge described and thus merits the award of that grade. GCSE examinations, therefore, consist of some elements of criterion-referenced assessment and some standards-referenced assessment.

2.2.4 Validity and Reliability of Assessments

Assessments such as a GCSE in economics and business must be trusted by all stakeholders to be an accurate representation of the measure of learning and knowledge and skills shown. Validity is the extent to which an assessment is an accurate measure of what students know and can do in the context of the aims and objectives of the assessment (Ofqual 2014a). Validity can be viewed as the extent to which the assessment measures what it was designed to measure and whether the questions/tasks/items measure what the assessment setter wanted them to assess (Killen, 2005, Wragg, 2001).

Reliability refers to the consistency of the assessment as a measure of learning, knowledge and skills (Ofqual, 2014b, Wragg, 2001), and how trustworthy, predictable and consistent the assessment is (Killen, 2005). If two students of similar ability, knowledge and skills following a GCSE programme of study in economics and business took an assessment, the outcome for both should be similar. A third party using these outcomes could be confident in the results (Black and Wiliam, 2006). For a GCSE in economics and business to be deemed reliable, the implication would be that a student taking a different version of the assessment (for example with different awarding organisations), or on a different day, or if marked by a different examiner, would receive the same or similar result (Ofqual, 2014b). If assessment results are inconsistent then the results or outcomes of the assessment would have little meaning and imply that decisions made on the basis of the results, such as whether a student is capable of moving on from GCSE study of economics and business to study A-level economics and business, would be compromised (Killen, 2005).

2.2.4.1 Characteristics of Validity

The validity of an assessment can be characterised in a number of different ways.

2.2.4.1.1 Content Validity

Any assessment will include subject specific content which is outlined in the specification. Content validity provides an indication of how relevant the content of an assessment is and the extent to which the assessment reflects the expected content of the course of study (Killen, 2005, Wragg, 2001). In economics and business the nature of the subject matter makes a focus on content validity particularly important. The specification for economics and business might indicate that students should understand how markets work and be able to work with the market model of supply and demand. Questions on supply and demand might be based on an expectation that students make some basic arithmetical calculations and have an understanding of mathematical concepts such as ratios and slopes. If questions are not framed carefully, the student might be being assessed not on the economics and business of the market model but on their arithmetic and mathematical skills and this would compromise content validity.

2.2.4.1.2 Concurrent Validity

There are different ways in which knowledge, understanding, learning and abilities can be assessed. In GCSE economics and business qualifications a variety of assessment instruments are used including objective test questions such as multiple choice questions (MCQs), alternate response questions (true/false), completion questions, matching questions, short answer

questions, structured questions and more extended response questions. One awarding body may use MCQs to assess understanding of costs and revenues in a firm whereas another might use short answer questions to assess the same content. Concurrent validity is the extent to which the form of assessment being used gives similar results compared to other ways of assessing the same knowledge, skills and understanding (Wragg, 2001).

2.2.4.1.3 Construct Validity

A construct is a theoretical conceptual framework, such as reasoning, ability, self-esteem, motivation or cognitive style, which represent a stand-alone theory to describe human characteristics, behaviours or groups of characteristics (Killen, 2005, Rayner, (personal correspondence, 2014). Construct validity is the ability of an assessment to measure the theoretical construct being measured, which forms part of the programme of study and whether appropriate inferences can be drawn from the results. For example, assume an assessment question in a GCSE economics and business examination is written to assess the student's understanding of motivation in human resources management. The question presents the student with a case study followed by a series of questions. It is possible that rather than assessing the student's understanding of motivation, the assessment inadvertently measures the student's reading or comprehension skills. The language used in the case study could be of a level which provides a challenge to students' reading abilities and so is assessing an irrelevant construct. It is equally possible that the way the actual question itself is phrased could challenge students' reading abilities and so examiners have to be mindful of what can be expected of the 'average 16-year old' when composing questions at GCSE level.

2.2.4.1.4 Predictive Validity

Predictive validity is the extent to which an assessment can be used to make predictions about future performance. The results of GCSE examinations are viewed and judged by many different people and groups, not only in the aftermath of the examination itself but in later life. Increasingly, universities in the UK expect to see 'good' GCSE grades to support A-level predicted grades in judging whether to make offers to students implying that 'good' GCSEs are indicative of consistency of approach and application by a student to their learning. Many employers ask for GCSE results to be included in job applications many years after they have been taken and when the applicant has gained considerable employment experience. Internally, schools use the results of summative assessments during the programme of study, such as mock exams, as a predictor of performance. It follows that to have a high degree of predictive validity the correlation between, for example, GCSE results and future A-level or university degree outcomes have to be statistically significant.

2.2.4.1.5 Issues of Validity

It is unlikely that any assessment can ever fully meet the previously stated and traditional categories of validity. Killen (2005) points out that questions can be linked to the curriculum thus satisfying content validity, test the appropriate conceptual understanding required (construct validity), match with scores on other similar tests (concurrent validity) and give an appropriate indication as to how the student might perform in other tests in the future (predictive validity) but at the same time have other aspects which it also tests which students may not be confident with or fully understand. For example, in an economics and business exam, questions which include arithmetic operations or the comprehension of a case study, students might not perform as well as they might even though the test meets all validity criteria. Killen (2005) concludes that it is inappropriate, therefore, to say that a test item or assessment task is valid in any absolute sense. Interpretations of test scores are a matter of judgement and inferences made can be valid or invalid (Killen, 2005). Wiliam (2000, p118) perhaps sums the issue up appropriately when he notes "validity is not a property of a test at all, but a property of the conclusions that we draw on the basis of the test results".

Phase I of this thesis reviews the performance of a range of students in GCSE examinations through content analysis. It is not clear from the research

conducted that any absolute inference can be drawn from the scores given to students by examiners in determining whether a student has begun to develop deep understanding of economics and business. Killen (2005) notes that low scores on a test might mean that students did not perform on that particular set of items successfully but that an inference that the student does not understand the content or is incompetent cannot be drawn. Other factors may have contributed to the weak performance on the test such as anxiety, fatigue, inattention on the particular day the test was taken, language proficiency, lack of prior knowledge, low motivation and so on. Killen (2005) summarises that it is not possible to validate a test, only the inferences drawn from the results. The challenge, therefore, is to produce assessments which generate evidence from which valid inferences may be drawn about learning (Killen, 2005).

It follows that the purpose of the assessment has to be clearly defined and understood by all those involved. In the context of GCSE, this means the regulator, awarding bodies, senior examiners, awarders, teachers and students. The aims and objectives of GCSE qualifications can be somewhat generalised and not easy to assess and if this is the case then inferences about performance are subject to question. The confusion between aims and objectives of the qualification on the one hand and the focus on generic assessment objectives on the other in GCSE qualifications makes the inferences that can be drawn subject to similar confusion. Teachers' approaches to the way they construct their teaching and learning programme will be influenced in part by the final assessment, an issue explored in more detail in Phase III of this thesis. How a teacher's belief about what it means to understand or think in the subject will have an effect on the way they influence the understanding of students and how they try to help learners understand. Given the current way in which GCSEs are regulated and developed, it is not clear that there is a coherent and consistent understanding amongst stakeholders about what it means to 'think in the subject' or even whether this is one of the aims of the assessment.

2.2.5 Reliability

If inferences are to be drawn from an assessment then the extent to which there can be confidence in the results is important. Reliability goes hand in hand with validity; one cannot have reliability without validity, for example, it is possible for an assessment to fail to measure what it is supposed to measure in a consistent (reliable) way (Wragg, 2001). In considering reliability thought has to be given to whether the assessment is worthwhile - what is worth assessing and measuring? (Wragg, 2001). The feasibility of an assessment also has to be taken into account (Wragg, 2001), which includes the time and resources that can be devoted to the assessment, how it is to be marked, how markers are standardised, how awarding of grades is managed, the quality of mark schemes used and the time and resources needed to teach the course or content which is being assessed.

Black and Wiliam (2006) highlight a number of issues which can be taken into account when considering and seeking to improve reliability in assessments. They note that if students take the same test on a number of occasions, the average score of all the results would represent a 'true score' - a good indicator of the measure. Analysis of individual components (items or questions) can be conducted to see the extent to which students with high overall scores have high scores on these items and whether students with low overall scores have low scores on items. If this is the case then the item can be said to have high discrimination and if all items have high discrimination then there is a degree of consistency which exists and so the assessment can be viewed as being reliable. The measure of the internal consistency between items is referred to as the reliability coefficient. A reliability coefficient of 1 means that there are no errors in the test and that it is perfectly reliable; a coefficient of 0 means that individual scores are all error and the test is completely random.

Black and Wiliam (2006) go on to note the trade-offs that have to be considered in increasing the reliability coefficient of an assessment. Assume that the reliability coefficient is 0.75 and the standard deviation (σ), 7.5, this means that for 68% of candidates the true score will be within one standard deviation (7.5) of their given score. 96% of candidates' true scores will be within two standard deviations (15) of their given scores and 4% of candidates will be at least =>15 marks away from their given scores. Black and Wiliam (2006) note that for a class of 30 students, such a scenario would mean that at least one student's score will be wrong by 15. Clearly this can be regarded as a serious question over the reliability of the assessment.

To increase reliability, it is possible to ask additional questions on fewer topics or make the test longer and ask more questions on all topics. The typical GCSE examination in economics and business is around 90 minutes long and may include anything up to 30 questions. If the reliability of a test is to be increased from 0.75 to 0.85, Black and Wiliam (2006) suggest that the test needs to be 1.9 times as long which would mean that the number of students misclassified would fall by around 8 - 9%. Wragg's point about the feasibility in relation to reliability (Wragg, 2001) becomes more relevant when considering ways of increasing reliability given Black and Wiliam's observations. Few teachers would be willing to see GCSE assessments increased in time significantly and the resourcing that all stakeholders would need to devote to the process would be considerable. It could be possible to reduce the scope of the content in the syllabus but successive revisions of GCSEs have failed to agree on any meaningful reduction in content and in conversation with some teachers there is a view that more content equates with greater demand and more rigour. Increasing reliability through extending the length of examinations and/or reducing content seems to be unlikely in the current political and educational climate. Re-conceptualising the focus of the assessment to align the understanding and focus of stakeholders is necessary and is a recommendation provided at the end of this thesis and outlined in the next section.

2.2.6 Taxonomies and Educational Objectives

The difference between the generic aims and objectives given by awarding bodies for GCSE economics and business qualifications and the generic

assessment criteria or objectives leads us to review how the focus on assessment objectives has arisen, the nature of assessment objectives and how learning and understanding can be assessed in different ways. Jephcote and Abbott (2005), note that Bloom's Taxonomy is pervasive in economics and business specifications and provides the basis for the learning, the design of modes of assessment and a rationale for the strategies for teaching and learning. This section of the literature review will look in more detail at Bloom's Taxonomy, revisions of the Taxonomy in the light of new research and an alternative taxonomy, the Structure of Observed Learning Outcomes (SOLO) Taxonomy (Biggs and Collis, 1982).

2.2.6.1 Bloom's Taxonomy

GCSE examinations are but one example where the pervasiveness of Bloom's Taxonomy exists. To give it its full title, the Taxonomy of Educational Objectives, provides a clue to its substance. Tyler (1949) noted that any educational programme had to have some goals and referred to these goals as 'educational objectives'. He went on to define educational objectives in terms of the behaviour to be developed and the context in which the behaviour operates. It follows that objectives in the context of assessment can represent desired behaviours. Davies and Brant (2006) define assessment objectives as generalised statements of capability which describe the type of information processing that the learner will demonstrate in relation to the subject content, and details the outcomes to be assessed.

GCSE specifications in economics and business define the types of outcome in language that is directly taken from Bloom et al's work and is a testament to the enduring popularity and importance of the work of Bloom and his colleagues. Hook (2012) notes that Bloom's work was a theory of knowledge and not a theory of assessment. Indeed, knowledge plays a fundamental role in the taxonomy.

Bloom et al (1956) describe a taxonomy as a classification according to natural relationships and a taxonomy of educational objectives as the classification of the goals of an educational system. Bloom's Taxonomy has become a catch-all for the goals of assessment systems and a guide to the development of assessment and teaching and learning programmes. Bloom et al (1956), ask the question 'what is to be classified?' and suggest that the basis of the classification could be teacher behaviour, instructional methods or intended student behaviour. It was decided that the focus of the taxonomy would be on classifying student behaviours which represent the intended outcomes of the educational process. This intended behaviour can be described in different ways. Bloom et al (1956) focused on a hierarchy of skills and abilities but it is equally acceptable to describe intended behaviours as the ability to think in the subject. The key point being that the intended behaviours defined will influence the way in which the assessment is constructed in order to test the intended behaviours and in turn the teaching and learning programme.

An important feature of Bloom et al's Taxonomy is that it represents a hierarchical structure representative of an order of different classes of objectives. Objectives of one class make use of and are built on behaviours found in preceding classes (Bloom, et al, 1956). In the way in which these skills are represented as assessment objectives in GCSE examinations, could it be implied that if a student can analyse or evaluate one piece of knowledge or content (for example, analysing or evaluating factors affecting the profitability of firms) that these skills are transferable to other pieces of content (for example, being able to also analyse or evaluate the factors affecting exchange rate movements)? Davies and Brant (2006) note that it is not possible to draw such an implication from Bloom's Taxonomy because analysis or evaluation is dependent on knowledge and understanding acquired particular to a context, a central assumption of the hierarchical nature of the Taxonomy. Davies and Brant (2006) argue that what the use of Bloom's Taxonomy in GCSE and other public examinations does imply is that the set of assessment objectives is applied to each part of the syllabus or specification content. A student might be able to demonstrate one or more skills in relation to one part of the subject content but not to another. It is unlikely that any examination could cover all subject content which may mean that some

students are disadvantaged given the different focus of the knowledge base (Davies and Brant, 2006).

The next section outlines the key features of the six major objectives of Bloom's Taxonomy (Bloom et al, 1956)

1. Knowledge:

Knowledge is classified in a variety of different ways. Knowledge of:

- specifics the recall of specific and isolable bits of information
- terminology such as the key terms in economics and business
- specific facts what is known in economics and business, for example, and which can be relied upon
- ways and means of dealing with specifics, of organising, judging, studying and criticising ideas and phenomena
- conventions, for example, how economists and business people represent concepts, ideas, use models and so on.
- trends and sequences and the relationships and processes which change over time, for example, trends in business activity or the impact of social media on business activity
- classifications and categories, for example, the different forms of business ownership or the classification of costs of production
- criteria, for example, how judgements are made on the success of a business or economic policy
- methodology, inquiries and techniques used in a discipline such as economics
- universals and abstractions in a field, the ideas schemes and patterns by which phenomena are organised and how a field can be described by structures, theories and generalisations
- principles and generalisations, for example, the way in which markets, individuals and firms behave in an economy

• theories and structures such as the theory of market structures, the behaviour of firms, how the labour market works and so on.

2. Comprehension

Knowledge has to be understood which implies making some use of the material or ideas. Comprehension involves an element of understanding to interpret the various types of knowledge, to translate and extrapolate, which involves making inferences, estimates or predictions. Comprehension or understanding is invariably included in the same sentence as knowledge in GCSE assessment objectives but Bloom et al (1956) noted that 'understanding' can mean different things to different people. Comprehending knowledge, internalising knowledge, grasping the core or the essence of knowledge could all be interpreted as representing 'understanding'. However comprehension is defined, Bloom et al (1956) make it clear that it involves recognising major ideas and inter-relationships contained in knowledge and going beyond mere repetition or rephrasing of knowledge.

3. Application

Comprehension is a prerequisite of application. This implies that knowledge can be used through appropriate abstraction and the way in which the abstraction is used, without prompting, in new situations and contexts. In GCSE economics and business examinations, application is interpreted as being able to carry over learning into new contexts where the knowledge being assessed is abstracted so that it is specific to that particular context. For example, in business, just-in-time (JIT) is a method of stock control and whilst many businesses carry stock, JIT is not an appropriate method in all these cases. The ability of the student to discriminate and use a method such as JIT in appropriate business contexts (and equally to know when it is not appropriate) is indicative of the ability to apply knowledge. Typically, examiners might look at a student's answer and ask themselves the question, 'could the way the student is applying the knowledge be generalised to any business or economic situation?' If the answer is 'yes' then it is likely that the student is not applying knowledge appropriately.

4. Analysis

Analysis is the breaking down of knowledge or information into its constituent parts in order to convey sense or meaning. The individual parts may have some interrelationship and the ability of the student to detect these relationships and to show how they are organised is characteristic of the ability to analyse. Bloom et al (1956) acknowledge that aspects of the ability to analyse are not only dependent on the ability to comprehend and apply knowledge but also overlap with evaluation. Examples given by Bloom et al (1956) serve to highlight this overlap. They refer to the ability to distinguish fact from opinion/hypothesis, identifying conclusions and supporting statements, distinguishing relevant from extraneous material, seeing how one idea relates to another, what unstated assumptions might exist and distinguishing between dominant and subordinate ideas as characteristic of analysis. However, to do some of these things a student is going to have to make a judgement which could be argued to be characteristic of evaluation. The acknowledgement that there are 'not entirely clear lines' that can be drawn between analysis and evaluation is a problem which faces examinations regulators, awarding organisations and examiners. In the revision of GCSEs in England and Wales in 2009, analysis and evaluation were included as one assessment objective (AO3). Producing mark schemes which appropriately reflect the different abilities of analysis and evaluation is a skill in itself and presents challenges to all those involved. This is a feature of the results and discussion in this thesis.

Bloom, et al (1956), distinguish between three types of analysis:

- a) The analysis of elements which requires students to identify both the explicit elements of what Bloom et al (1956) refers to as a 'communication' and others which are not so explicitly stated.
- b) Once elements are identified, relationships between the elements have to be determined and some judgement made on which elements are important or relevant and which are not. It may also be necessary for the student to identify elements which are related through cause and

effect, which elements are assumptions or associated with assumptions and how ideas are linked together. This is the analysis of relationships.

c) The analysis of organisational principles requires the student to look at the structure and organisation of a communication such as the intention of the author of a piece of information, the point of view, degree of bias, perspective being portrayed and the techniques used to present an argument.

5. Synthesis

Synthesis refers to the putting together of elements and parts to form a coherent whole. Students who synthesise are able to identify patterns and structures in a body of information or given knowledge which is not immediately present. Synthesis requires a degree of creativity from the student and implies that the student is actively, not passively, involved in the material and is able to exercise some degree of independent thought and personal expression.

In assessment this can bring some problems because the results of the synthesis provided by a student might not match with those of the assessor and unless the mark scheme is very flexible and the examination system used encourages this sort of approach by students, the skill may not be recognised. The process of synthesis might also be the product of time rather than something that can be assessed at a particular point in time such as in the confines of an examination. A student's examination answer may not be an accurate reflection, therefore, of ability in this skill area. It is perhaps not surprising that for this reason, and because of the difficulties in passing judgement and the standards to be applied to an end product by a student, that synthesis is not explicitly included in the assessment objectives in English examination systems including those for GCSE.

6. Evaluation

Evaluation is one of the most difficult and contentious of the skill areas outlined in Bloom's Taxonomy. Bloom et al (1956) define evaluation as making

judgements about the value, for some purpose, of ideas, works, solutions, methods and materials. Making judgements or placing value on ideas etc. could be quantitative or qualitative.

Bloom et al (1956) see evaluation as the last stage in a complex process which involves a combination of all other behaviours in the Taxonomy but not necessarily the last step in thinking or problem solving. Instead, evaluation might be the prelude to the acquisition of new knowledge, a new attempt at comprehension or application or a new analysis and synthesis. In this respect, Bloom et al reflect the cycle of learning put forward by Kolb (1984).

Evaluation presents challenges to those constructing mark schemes for examinations because there can be different levels of evaluation. Bloom et al (1956) note that judgements made by quick decisions might be more properly referred to as opinions. In arriving at more sophisticated levels of evaluation, assessors might be looking for support to be provided for the judgements made drawn from the context, from the student's own knowledge or from material presented to the student.

Davies and Brant (2006) also point out that there needs to be clarity on the definition of evaluation used - the evaluation of ideas, evidence and argument, which Bloom et al (1956) propose, or the evaluation of decisions, opinions, or the options of a decision maker, which is a common aspect of economics and business. The language of assessment objectives provided by awarding organisations, according to Davies and Brant (2006), can often be ambiguous in this respect as to which or whether either or both of these interpretations are intended.

The blurred lines between analysis and evaluation is highlighted by the fact that distinguishing between fact and opinion, which may involve a judgement, is considered by Bloom et al (1956) to be analysis, whereas Davies and Brant (2006) point out that this could be legitimately considered as part of evaluation in assessment objectives for examinations.

Bloom's Taxonomy has been subject to revision since its publication and two particular revisions will be considered. The first is by Anderson et al (2001)

and the second by Marzano and Kendall (2007).

The revision by Anderson, et al (2001) included as one of its authors David Krathwohl who was also one of the original researchers working with Benjamin Bloom in 1956. Anderson et al's (2001) revision is based on a new two-dimensional framework of cognitive process and knowledge. They identify six categories of the cognitive process:

- Remember
- Understand
- Apply
- Analyse
- Evaluate
- Create

It is assumed that each category of the cognitive process is more complex than the previous one so that the skill of applying, for example, is more complex than understanding. The knowledge dimension contains four categories which provide a continuum from the concrete to the abstract:

- Factual
- Conceptual
- Procedural
- Metacognitive

Anderson et al (2001) note that the statement of objectives contains both a noun and a verb and describes the intended cognitive process and the knowledge (content) to be acquired or constructed. Cognitive process replaces the term 'behaviour' because of its associations with behaviourism which Anderson et al (2001) were keen to distance themselves from.

The revised framework is a hierarchy with the six categories of the cognitive dimension ordered in increasing complexity. Anderson et al, (2001) note that the original Taxonomy assumed the categories to be cumulatively hierarchical Page | 58

where mastery of a prior skill was required before mastery of a more complex category. Research, they suggest, pointed to some evidence for this being the case for comprehension, application and analysis but was much weaker for the latter two categories of synthesis and evaluation.

The revised Taxonomy allows the six categories to overlap on a scale of judged complexity so that understanding, for example, is much broader than just comprehension and explaining something can exemplify a kind of understanding but is much more complex than simply applying at its lowest level. Create was judged to be a more complex category than evaluate but problem solving was not included as it was deemed to call for cognitive processes in several categories.

The focus of national examinations such as GCSE on assessment objectives which are based on Bloom's Taxonomy necessitates a detailed description of its elements. The pervasiveness influence of the taxonomy shapes the nature of the assessment, influences the way in which teachers view the assessment and the way they plan their teaching and learning activities. If there is to be a change in the approach of teachers then the assessment objectives must change and that means a move away from Bloom's Taxonomy as the basis of assessment to assessment objectives which more appropriately define and allow measurement of desired learning outcomes which focus on deep learning and thinking in the subject. An attempt to define revised assessment objectives is provided in Chapter 7.

2.2.6.2 The Cognitive Process Dimension

The cognitive process dimension highlights the differences between Anderson et al's (2001) approach and that of the original Bloom Taxonomy (1956). The dimension consists of the following:

1. Remember

Remembering consists of three sub-levels. Remembering itself could be where instruction promotes the retention and recall of knowledge and the retrieval could be of factual, conceptual, procedural or metacognitive knowledge or a

combination of these. Remembering is deemed as essential for meaningful learning. Remembering might also include recognising - the retrieval of relevant knowledge to compare with presented information (also referred to as identifying) and recalling which relates to the retrieval of knowledge from long term memory when prompted.

2. Understand

Understanding becomes relevant when the goal of instruction is to promote transfer. Understanding, suggests Anderson, et al (2001), means constructing meaning from instructional messages and occurs when new knowledge is connected to prior knowledge. Understanding requires various cognitive processes such as:

- interpreting the conversion of one representational form to another;
- exemplifying giving a specific example or instance which requires an appreciation of the defining features of the general concept or principle;
- classifying the recognition that something belongs to a certain category requiring relevant features or patterns to be identified which fit to the defining features of the specific principle, concept or instance;
- summarising the suggestion of a single statement that represents presented information or abstracts a general theme;
- inferring finding patterns;
- comparing detecting similarities and differences between two or more objects, events, ideas, problems or situations
- explaining the constructing and using of cause and effect to show how changes in one part of a system or model affects changes in another.

3. Apply

Apply makes use of two cognitive procedures, one to execute a task (when the task is familiar) and another to implement (when the task is unfamiliar).

4. Analyse

Analyse involves breaking down material into its constituent parts and determining how the parts are related to each other and to an overall structure. The skill of analysing includes differentiating which requires an ability to distinguish parts of a whole structure in terms of relevance and importance and discriminating between relevant and irrelevant, important and unimportant. Whilst Anderson et al (2001) characterise these as elements of analysis, there seems to be some requirement for students to make judgements in distinguishing between relevant and irrelevant, important and unimportant. This may be indicative of the suggestion that the categories do overlap but the extent of the overlap in this case could be argued to be considerable. A similar contradiction arises in the sub-category of attributing which is where a student is able to ascertain the points of view, biases, value or intentions underlying a communication which Anderson et al (2001) also refer to as deconstructing. Again, the ability to determine biases, values or intentions could require some element of judgement which would indicate an overlap with evaluation. Finally, Anderson et al (2001) suggest organising as a sub-category of analysis which involves identifying elements of a communication or situation and recognising how they fit together in a structure.

5. Evaluate

Anderson et al (2001) define evaluation as making judgements based on criteria and standards with the criteria tending to be quality, effectiveness, efficiency and consistency. The criteria could be determined by the students themselves or by others for students to judge. Anderson et al (2001) note that not all judgements are evaluative.

Evaluation includes the sub-categories of testing for internal consistencies or fallacies in an operation or product, for example, whether a conclusion follows from its premises or whether data supports or rejects a hypothesis and whether material throws up contradictions which are referred to as 'checking'. Critiquing involves judging a product or operation based on externally imposed criteria and standards. Students, for example, may be

required to note the positive and negative features of an issue and arrive at a conclusion based on these (and potentially other) features. This, it is argued, is at the core of critical thinking.

6. Create

Create is defined as the putting together of elements to form a coherent or functional whole. Anderson et al's (2001) definition here seems to echo the Bloom category of synthesis but in this revised taxonomy is elevated to a more complex process than evaluation. Create is likely to require aspects of the other five cognitive processes but not necessarily in the order in which they are listed in Anderson et al's revised taxonomy. They break create down into three sub-categories, generating, planning and producing. Generating involves representing the problem and arriving at alternatives or hypotheses to meet certain criteria. When transcending the boundaries or constraints of prior knowledge and existing theories, generating involves divergent thinking (arriving at various possibilities) and forms the core of creative thinking. Planning is devising a solution that meets a problem's criteria and developing a solution which Anderson et al (2001) also refer to as 'designing'. Finally, producing is the carrying out of a plan for solving a given problem that meets certain specifications and can require the coordination of the four types of knowledge also called 'constructing'.

2.2.6.3 The Knowledge Dimension

The knowledge dimension consists of four main types, factual, conceptual, procedural and metacognitive. Factual knowledge is the factual knowledge of a discipline and is also referred to as declarative knowledge – knowing what (Boulton-Lewis, 1995) is. Such knowledge might include the terminology and specific details and elements which make up the discipline (Anderson et al, 2001). Conceptual knowledge has three sub-types, the knowledge of classifications and categories, the knowledge of principles and generalisations and the knowledge of theories, models and structures. Procedural knowledge is the 'knowing how' (Biggs and Tang, 2007), and includes the knowledge of subject specific skills and algorithms, subject specific techniques and

methods and the knowledge of when to use the appropriate methods using some criteria. (Anderson et al, 2001). Procedural knowledge allows for the purposeful manipulation of declarative knowledge to undertake a task, solve a problem, make decisions and show understanding (Boulton-Lewis, 1995). Metacognitive knowledge is the knowledge about cognitive tasks and selfknowledge. It includes conditional knowledge, knowing when to use certain procedures and be able to access appropriate content referred to as the 'knowing how and why' (Boulton-Lewis, 1995).

The Anderson et al revision of Bloom does contain elements of the same thinking which defined the original taxonomy and whilst the blurring of the categories in the cognitive domain does recognise some of the advances in research into cognitive psychology, it does not go far enough and this is what prompted Marzano and Kendall (2007) to develop a further revision. Marzano and Kendall's main criticism of both Bloom's Taxonomy and the Anderson et al (2001) revision is that they both over simplify the nature of thought and its relationship to learning. Their argument is that both taxonomies attempt to use degrees of difficulty as the basis for the different levels but psychology suggests complex processes can be learned with little or no conscious effort on the part of the student. Phase I of the research highlights the way in which students can be taught processes and ways of answering questions which give a veneer of understanding and gain marks according to the mark schemes.

Marzano and Kendall (2007) suggest that the difficulty of a mental process is a function of two factors, the inherent complexity of the process and the level of familiarity of the student with the process. Mark schemes for GCSE exams, for example, give teachers a significant amount of guidance about the way in which the marks will be awarded and what processes the student needs to follow and this point is taken up further in Phase I of the research.

The Marzano and Kendall (2007) taxonomy involves the 'self-system' whereby judgements are made by the student as to whether it is worth engaging in the task at hand and is reflective of the motivation of the student. The self-system involves the student perceiving the extent to which the task at hand is

more important than other tasks. The metacognitive system sets goals for the task and designs strategies for accomplishing a goal. The cognitive system is responsible for effective processing of information to complete the task set, for example, analytic operations, making inferences, comparing, classifying and so on. The success of the task is dependent on the knowledge the individual has about the task and the amount of prior knowledge. The assumption of prior knowledge can be overstated by teachers, especially where cross-curricular application of knowledge is concerned, for example the assumption that students who are doing GCSE maths will be able to transfer knowledge of ratios, percentages, averages and so on, to their study of economics and business.

The new taxonomy devised by Marzano and Kendall (2007) is a twodimensional model with three knowledge domains representing one dimension and six categories of mental processes representing the other dimension, summarised below:

The Knowledge Domains:

- The Domain of Information
- The Domain of Mental Procedures
- The Domain of Psychomotor Procedures

The Mental Processes:

- Level 1 Retrieval (Cognitive system)
- Level 2 Comprehension (Cognitive system)
- Level 3 Analysis (Cognitive system)
- Level 4 Knowledge utilisation (Cognitive system)
- Level 5 Metacognition
- Level 6 Self-system Thinking

The new taxonomy separates the various types of knowledge from the mental processes that operate on them. Marzano and Kendall (2007) justify this

approach by noting that Bloom's Taxonomy mixes various forms of knowledge with mental operations and state that mental operations interact differently with knowledge domains. The knowledge dimension is organised into three categories, information, mental procedures and psychomotor procedures. Information in economics and business would include markets, the behaviour of firms, the macroeconomy, human resources management, marketing, finance and operations management, for example. Mental procedures would include reading graphs, using maths to calculate accounting ratios, reading profit and loss accounts and balance sheets and so on. In economics and business there are few psychomotor skills employed unlike, for example, physical education where performance at different skills is a requirement of the GCSE assessment.

Marzano and Kendall (2007) identify three distinct phases in the acquisition of mental procedures, a cognitive phase where the learner can verbalise or describe the process, the associative stage where the performance of procedure is smoothed out and the autonomous stage where the process is refined and becomes automatic. Mental procedures are categorised into two, those which are automatic and performed with little conscious thought and those which must be controlled. The former may include skills, tactics and algorithms used by the student and in economics and business might involve reading supply and demand diagrams, knowing axes, reading the slopes of curves and recognising the equilibrium price and output but not necessarily in any specific order of identification. The latter may be referred to as macroprocedures such as students answering questions on a particular topic but arriving at very different outcomes.

The new taxonomy devised by Marzano and Kendall (2007) has some similarities with both the Bloom and Anderson et al taxonomies in that it remains hierarchical. Its hierarchy, however, is based on a flow of processing rather than levels of complexity. If a student considers a task is important then the metacognitive system establishes learning goals and plans to carry out the goals. The cognitive system is responsible for the process that enables the student to achieve the learning goals.

The new taxonomy has implications for assessment. To develop assessments which measure competence in general topics, Marzano and Kendall (2007) argue that test makers construct items which sample information or skills within the topic. Objectives specify the content and behaviours associated with the content that is to be the focus of the instruction. The instructional objective includes three elements:

- Performance which specifies what a learner will be able to do
- Conditions under which the performance is to occur
- A description of the criterion to judge an acceptable performance.

The new taxonomy adopts the approach that the student 'will be able to...' and then states the mental processes such as retrieve, comprehend, analyse, utilise and so on, followed by the object, the informational knowledge and forms of mental procedure. In awarding organisations' specifications this approach equates to the overall aims and objectives of the qualification. Such overall aims and objectives outline the knowledge and understanding and skills required or expected.

2.2.6.3.1 Summary

Whilst there have been revisions of Bloom's Taxonomy, many of the original elements of Bloom persist in the assessment objectives in GCSE specifications. In most cases the assessment objectives are presented in very brief statements which leave much to be interpreted by the teacher, the student and examiners. Revisions have attempted to update the taxonomy in the light of new research in psychology but these revisions have not had a significant influence on the design of assessment objectives in GCSE qualifications (or indeed most other national qualifications). The focus of the assessment objectives based on a Bloom type taxonomy is very much on what a student can do and does not focus sufficiently on the way in which a student thinks. Bell (1979), believed that Bloom's Taxonomy did not take into account the 'complete encompassing of human activity in an experience in the

planning'. Bell devised an experiential taxonomy which more appropriately reflects thinking and is perhaps more closely reflective of the idea of 'thinking in a subject' than Bloom or any of the revisions have managed to achieve.

Bell's Taxonomy (Bell, 1979) has five categories:

- Exposure the consciousness of an experience which consists of a sensory stimulus to an experience, a response to the stimulus and a readiness which anticipates participation.
- Participation is the decision to become involved and consists of representation - the reproduction mentally and/or physically of an existing mental image of the experience and modification, and the use of past experiences to grow and develop the new experience.
- Identification is the coming together of learner and objective/idea in an emotional and intellectual context for the achievement of the objective. If reinforced the experience is modified and repeated; if emotional, the learner becomes identified with the experience and will come to translate to an intellectual commitment which then leads to the learner sharing the experience with others.
- Internalisation leads to the learner's behaviour changing as the experience changes their lifestyle. Expansion occurs when the learner's attitudes and activities change and when the change becomes something more than temporary, the learner moves to intrinsic where the changed life-style is characterised more consistently.
- Dissemination consists of two elements, informational, where the learner informs and seeks to stimulate others, and homiletic where the learner sees the experience as imperative for others to also be exposed to.

Bell's Taxonomy (Bell, 1979) represents a way in which students may become involved in a subject area and becomes a passionate advocate for the subject experiencing a transformed understanding which affects their life but is difficult to translate into a meaningful assessment which would be valid and reliable. The assessment of ways of thinking which focus on qualitative outcomes rather than a hierarchy of skills which present its own challenges in assessment, represents an alternative approach to assessments based on Bloom's Taxonomy. The next section will explore one such approach, the Structure of Observed Learning Outcome (SOLO), (Biggs and Collis, 1982).

2.2.7 The Structure of Observed Learning Outcome (SOLO)

According to Biggs and Collis (1982) for meaningful learning to occur an educator needs to evaluate how much learning has taken place (quantitative) and how well the learning has been internalised (qualitative). The authors argue that the quantitative is relatively easy to measure but that the qualitative is far more subjective and the terms not always spelled out to students. SOLO aims to present a criterion-referenced measure of the quality of learning. Bloom's Taxonomy, Biggs and Collis (1982) argue, is intended to guide the selection of items for a test rather than evaluate the quality of a student's response. This implies that in order to ensure adequate coverage of the assessment criteria, awarding organisations specify the weighting of the assessment objectives and questions have to be devised carefully to match the relative weightings. In theory, the student cannot get high marks or grades through offering a large number of low-level responses. It is difficult to apply the taxonomy to questions which require extended written responses. Mark schemes have to adopt a levels-of-response approach and it is not always the case that teachers and assessors agree on the split of the assessment objectives across the question.

The main difference between Bloom's Taxonomy and SOLO, according to Biggs and Collis (1982), is that Bloom levels are *a priori*, that is imposed in advance by the teacher whereas SOLO uses levels that 'arise naturally in the understanding of the material'. Qualitative assessments reward students in terms of the structural complexity of the outcome and is linked to a general classification of levels of information processing. SOLO is based on the belief that there are natural stages in the growth of any learning and skill and relies heavily on Piagetian developmental stages in thinking. However, the learning quality is also influenced by external factors to the student such as the quality of the instruction received and on internal factors to the student such as learner motivation, their developmental stage and the extent of prior knowledge, amongst other things.

Biggs and Collis (1982) outline the assumptions of what they refer to as 'stage theory'. These assumptions are that stages proceed in irreversible sequences and are stable with the learner thinking in a characteristic of the current stage and not earlier or later stages. If a student 'fails' at one task it is likely to predict that they will also be unable to understand some other, similar, tasks. For Biggs and Collis (1982) the frame of reference of instruction changes so that the focus is on learning quality and this becomes the point of departure rather than the developmental stage of the student. Their research found that the assumptions of stage theory do not hold. Across different subject areas students demonstrated different stages. Biggs and Collis (1982) make a distinction between the cognitive structure of the individual and the structure of the responses to a learning task, which they refer to as the structure of observed learning outcome. An individual's developmental stage can determine the upper limit of functioning but other factors such as motivation and the extent of prior learning would determine whether or not the individual functions at that level. The student is not, suggest Biggs and Collis (1982), jumping from one developmental stage to another but manifesting that they understand some tasks better than others or is more motivated to demonstrate the understanding than on another task.

The SOLO taxonomy consists of a series of levels which describe a particular performance at a particular time. The taxonomy is not meant to be a series of labels with which to tag people but describes responses. The levels can be deconstructed in terms of cue and response and the individual's reaction. There will be a capacity issue in terms of the extent to which the individual is able to think of several things at once. If capacity is higher then there is a greater chance of making a greater number of connections, relationships and interrelationships and this greater degree of complexity is characteristic of higher levels of the taxonomy.

A relating operation looks at the way in which the cue and response interrelate. Transduction refers to a 'guesstimate' response which is not based on an adequate logical basis of selection; instead the response is based on emotional or perceptual basics. Induction, by contrast, is the process of correctly drawing general conclusions from particular instances. At different levels there will be variations in the degree of consistency and closure. Quick closure will utilise limited data and lead to simple responses to the cue whereas at higher levels, more data will be marshalled.

The structure outlines the ways in which the cue and response are related and the degree of complexity of the relation. At the prestructural level, the cue is related to the response by an irrelevant feature. Unistructural sees the cue related by a single relevant feature to the response and multi-structural has several relevant features. The relational level sees relevant data and elements of the parts begin to be tied up the into a conceptual structure and where the parts are seen in relation to the whole. The last level, extended abstract begins to demonstrate a far greater degree of complexity and understanding. A response at this level includes a wider range of relevant data and interrelates and subsumes these under hypothetical abstract structures that enable deductions to apply to instances that were not included in the original data and as a result the response might allow alternative outcomes. SOLO allows for transitional levels where the individual may be 'feeling' for the next level but does not quite make it and may be marked by some confusion or inconsistency and loses the track of the argument.

Killen (2005) argues that the SOLO taxonomy is based on the idea that useful inferences on student understanding can be made by examining the way in which written or aural answers are structured and also how teaching and questions can be arranged to encourage different structures. Killen (2005) provides the following descriptions of the different levels in SOLO which

provide some further expansion of the summary above:

- Prestructural very little, if any understanding demonstrated; the student will tend to repeat the question, guess the answer and show little or no engagement with the assessment task. Irrelevant detail may be given as part of the response and any justification for the response is likely to be personal and subjective, for example, suggesting that the answer 'looks alright to me'.
- Unistructural simple, correct conclusions are drawn and based on only one relevant aspect of the data present. The student sees the need to provide a quick decision to reach closure quickly and disregards possible inconsistencies provided in the answer. The response might be a concrete, given fact and the student may not understand why they have arrived at an answer.
- Multi-structural response the use of two or more pieces of data given to reach a correct conclusion. The answer will contain elements consistent with the question but lacks integration.
- Relational response the answer will make use of an overall concept or principle to account for various pieces of given, isolated data, and is used to integrate multiple ideas and links. The answer is likely to attend to several issues simultaneously. In economics this might be exemplified where the student is able to deal with supply, demand, price, equilibrium and shifts and movements along curves in explaining changes in market activity. Connections are made between a principle and the representation of the principle but the understanding of the broad principle may be limited and responses are inconsistent across different contexts and the student might have problems in applying the broad principle to different contexts.
- Extended abstract answers go beyond what is taught and logical deduction is used to frame the answer using knowledge which may not have been given to apply to the situation at hand. The student may

have no compulsion to reach a definite conclusion or closure and may be more concerned with resolving inconsistencies. Such answers may contain the 'it depends' approach where the student is able to hypothesise and reflect on different approaches, situations and perspectives.

SOLO has been used in higher education and some attempts have also been made to apply the taxonomy to secondary education (Martin, 2011). Martin attempts to link the taxonomy to the teaching of science and provides a series of examples of questions which can target the different levels and what sort of responses might characterise the levels. The attempts could be argued to be too simplistic and inaccurate a representation of SOLO. For example, Martin (2011) claims that higher levels of SOLO *cause* deep learning. Higher levels of SOLO could be argued to be characteristic of deep learning but it is debatable whether the taxonomy can cause an approach to learning or an outcome of learning.

Hook (2013) cites a claim from John Hattie that Bloom's Taxonomy is used more by teachers than students and that students can be taught to understand SOLO. The same could be said, however, of Bloom's Taxonomy and many teachers will make students aware of the assessment objectives and use mark schemes to explain to students what is expected by the assessment objectives. Hook (2013) also cites Hattie's claims that scoring using SOLO can be more dependable (implying some confusion between points-based marking and levels of thinking which is characteristic of SOLO) and that scoring is 'relatively easy' using SOLO. The evidence in this thesis drawn from Phase II would suggest that this is not necessarily the case and that a degree of familiarity with the levels is necessary.

Chan, Tsui and Chan (2002) looked at the degree of inter-rater reliability between markers using SOLO. Inter-rater reliability, they found, was higher when more sub-levels were included but they also noted that the degree of ambiguity might also be influenced by the length of assignments. The sublevels could be as simple as low, moderate and high in each SOLO level. Chan
et al's (2002) research also showed that different markers have different interpretations of what skills mean or look like in student answers and that the student outcome can also be dependent on the quality of the question - a poor question aiming to target higher order skills of evaluation and analysis might be framed such that all the student is really required to do is offer an opinion which might be reflective of a lower SOLO level response to the cue.

One of the appealing aspects of SOLO in comparison to assessment objectives based on Bloom's Taxonomy is that it does focus on thinking. Boulton-Lewis (1998) attempted to use the SOLO taxonomy to develop and assess higher order thinking in higher education. The paper assumed that the purpose of higher education was to develop deep knowledge in students which was qualified by being described as a change in conceptions and worldview and learning how to think critically. Critical thinking is assumed to rest on deep rather than surface learning outcomes and hence deep approaches to learning. If the intention of higher education is to encourage deep learning then Boulton-Lewis (1998) notes that the sort of assessment students receive must be designed to not only test deep learning but also encourage it. It is noted that a major criticism of tests is that they tend to emphasise recall of declarative or procedural knowledge and in turn provides limited evidence of the level of understanding or the quality of thinking. Assessment procedures, it is argued, need to be carefully designed to encourage and develop aspects of learning that lead to higher order thinking. Bloom's Taxonomy, and the attempts to revise it, identifies a number of issues which raise questions about its efficacy in encouraging and developing deep learning. Boulton-Lewis (1995) argues that this is a desired objective of higher education but can this also be assumed to be the case at GCSE? Given that GCSE specifications are focused on assessment objectives derived from Bloom's Taxonomy, it can be suggested that the assessment of students at this level might focus more on declarative and procedural knowledge rather than the higher order skills that characterise thinking in the subject.

The extent to which these higher order skills are reflective of deep learning is

open to question. Student answers to questions are awarded marks in relation to the assessment objectives and the student is then awarded a grade which is in part determined by the total mark scored but also on a combination of the judgement of senior examiners and by reference to statistical analysis of previous cohorts. The grade descriptors (see Appendix I) are reflective of the assessment objectives but provide a more generic description of the knowledge and skills expected at different grades. The assessment objectives, based on Bloom's taxonomy (Bloom et al, 1956), for GCSE economics and business are:

- Assessment Objective 1: Recall, select and communicate their knowledge and understanding of concepts, issues and terminology.
- Assessment Objective 2: Apply skills, knowledge and understanding in a variety of contexts and in planning and carrying out investigations and tasks.
- Assessment Objective 3: Analyse and evaluate evidence, make reasoned judgements and present appropriate conclusions.

AO1 and AO2 are weighted at 35% and AO3 at 30%.

Whilst there are links between the grade description for grade A (see Appendix I) and the assessment objectives, the breaking down of the assessment in this way presents problems when assessing learning. The mark scheme for the examination guides the marker in interpreting student answers in relation to the assessment objectives. For a 12 mark question, for example, the mark scheme may award up to 2 marks for demonstration of knowledge, up to 2 for application and 8 for analysis and evaluation. This provides teachers with a guide to the ways in which they can coach their students to meet the requirements of the mark scheme. It can be argued that this does require some learning – learning how to interpret and exploit the mark scheme - but it does not necessarily focus on the quality of learning.

Students arrive at the examination knowing they will get certain types of questions and will have invariably been coached using past exam papers.

Exam papers, in turn, have to have some consistency of approach and questioning to be considered as maintaining standards over time. Teachers can explain to students that if they are able to make some reference to appropriate knowledge they will pick up at least 2 marks; if they can answer in context they can pick up a further 2 marks; assuming that they can construct even a basic argument - offer some points for and against and summarise with a conclusion albeit a very simplistic one - they are going to amass a number of marks, possibly 8 out of the 12 available.

Other parts of the paper will have questions that are a great deal more accessible and which students can be trained to provide answers to. A question such as: 'Using an example, explain the meaning of the term 'market segment'' may be targeted at knowledge and application and be worth 4 marks. Providing an appropriate answer to such a question can easily be coached to students of all abilities enabling them to access at least 3 of the 4 marks if not the full 4.

This approach to learning means that students are able to pick up marks and get into the lower 70% region. Such a percentage is likely to be rewarded with a grade A. The question can be asked about the extent to which the learning the student has demonstrated in this scenario matches the grade descriptor and how far it reflects qualitative understanding. It is approaches to learning and the characteristics of deep learning and thinking in the subject that is the subject of the next section.

2.3 Approaches to Learning

The assessment of learning raises questions about what sort of learning is being assessed. At GCSE, assessment objectives are based on Bloom's Taxonomy which influences the way in which teachers and students approach learning. Marton and Säljö (1976) suggest that learning can be seen as being an outcome where the number of correct answers to a test defines the learning. Marton and Säljö (1976) prefer to focus on the idea of *what* students learn rather than *how much* they learn. In so doing, Marton (1983) sees two different conceptions of learning; knowledge as ways of viewing reality and Page | 75 knowledge as a collection of right answers. Marton (1983) highlights the subjective nature of the understanding of learning in that a teacher will be looking at a learning task in a different way than the student or subject may be viewing it.

Marton and Säljö (1976) concern themselves with the individual meanings that students assign to the same learning task and content. The learning outcome could be seen as being correct and adequate by one member involved in the task but wholly inappropriate and wrong by another. If the student sees learning as securing a sufficient number of marks to get a particular grade, this may not equate to a definition of learning that sees it as a reshaping of the learners understanding of reality and their ability to tackle new tasks *using* the knowledge gained. Marton (1983), suggests that we have to go beyond what our perspective of learning might be and observe and seek to understand others' experience of learning, referred to as phenomenography. The phenomenographical research method developed by Marton and his colleagues at Gothenburg (Marton, 1981) attempts to recognise and highlight this different perspective.

Entwistle (1987), builds a model of approaches to learning that include deep, surface and strategic approaches but also includes motivation (a whole field in itself) and affect. The outcome is a heuristic model of learning where the learning is clearly focussed on the classroom rather than a more general definition of learning. The three approaches to learning have particular characteristics with the difference between surface and deep approaches to learning of critical importance.

Surface learning is characterised by memorisation, of the learning of facts or terminology without associated understanding which limits the student's ability to use these facts and terminology in different contexts and when new situations are presented. Text represents discrete elements and there is a limited attempt to make links between different elements. Concepts and theories may be presented but lack reflection and the meaning is missed (Entwistle, 1987). Surface approaches to learning may be appropriate in

certain situations - where questions simply require the student to give a definition and an example to gain 4 marks, for example.

Deep learning is characterized by an intention to understand the material and a willingness to interact with the subject content. Deep learning exhibits the characteristics of higher order skills in learning - analysis, synthesis and evaluation- and the student is able to see the whole yet relate concepts, methods and theories to each other. There is a capacity for relating existing knowledge with new knowledge and for relating concepts to everyday experience and to new contexts as they arise. Deep learning is also characteristic of a willingness to look beyond the subject to help build understanding, (Marton and Säljö, 1976; Entwistle, 1983; Ramsden, 1992; McCune, 1998; McCune and Entwistle, 2000; Entwistle, McCune and Hounsell, 2002, Biggs and Tang, 2007).

The relational and extended abstract levels of SOLO could be said to be reflective of deep learning whilst prestructural, unistructural and multistructural levels are more characteristic of surface approaches to learning. Strategic learning is characterized by students being clear about the outcomes of learning and adopting different approaches to learning to secure that outcome (Entwistle, 1987, Biggs, 1987). Strategic learners aim to achieve high grades and will organize their time and effort appropriately to achieve this end. When appropriate, surface approaches will be adopted, for example, where students know that certain questions or topics will only require rote learning (such as questions where students are required to give definitions) and in other cases the student may intend to understand the material where they believe it is required by the examiner in order to access the marks available. Strategic approaches to learning may well see students using past papers, published mark schemes and command words as a key part of their preparation for examinations to maximize the potential for securing marks and achieving the desired grade. It is not only the perspective of the student that is relevant with regard to strategic approaches to learning. Teachers judged by exam results may adopt teaching strategies and pedagogy

which encourage strategic approaches to learning. This implies that deep learning and thinking in the subject is not the primary focus of teaching and learning by both student and teacher when assessment is based on Bloom's Taxonomy. If the assessment was based on the SOLO taxonomy, a question arises as to whether the teaching and learning and approach to learning of both students and teachers would change and be more focused on deep learning?

Marton and Säljö (1984) suggest that deep approaches are not always the 'best' but that it is the only way in which full understanding of learning materials will be gained. This view is supported by Ramsden (1992).

Entwistle (2003) and colleagues looked at links between approaches to learning and the quality of the learning achieved. Entwistle (2003) identifies a conceptual framework indicating influences on student learning. It includes student approaches, student perceptions, the perceptions about the teaching and learning environment of the lecturer, course material and content, the design of the teaching and learning environment and existing knowledge and skills of the student.

This framework is also echoed in other studies on the topic. Laurillard (2002), discusses different approaches to teaching strategy and design; the link between teaching and student learning. Laurillard (2002) looks at three different approaches, Gagne's instructional design based on information processing theory, constructivist psychology and phenomenography. Laurillard (2002) dismisses the instructional design model citing its lack of empirical basis, whilst the constructivist approach and Biggs' (1996) constructive alignment offer some pointers to a bridge between what we know about student learning and what teachers should do as a consequence.

2.3.1 Thinking in the Subject

Assessment objectives can be seen as a primary driver in the approach to learning by both teacher and student. It has been argued that basing assessment on Bloom's Taxonomy may lead to atomised learning through which the student approaches learning in a strategic way. There can be confusion as to the aims of the assessment as a result and progression to the next stage of learning may be compromised because students do not develop habits of learning which lead to high levels of thinking but instead ever more strategic ways of securing the grades needed to move to the next stage of education. Expertise in the subject area is similarly compromised.

It was stated at the outset of this thesis that the definition of learning being used was one which involves a change in thinking and a change in the way the student sees and thinks about the world as a result of their studies. This can also be interpreted as being the ability to 'think in the subject'. Adopting this as a primary aim of the learning experience would require the assessment be designed to encourage the focus of teacher and student to be on this which in turn then influences the approach to teaching and learning and pedagogy.

The progressive nature of the educational system in England and Wales leads to students having to use more complex knowledge and face increasing demands throughout their education. Advanced level study, (generally the 16 - 19 age group), places a greater emphasis on the development of higher order skills than at GCSE, and at undergraduate level this is taken one step further and so on. This progressive system leads to a degree of specialisation in particular disciplines. As the student moves through, they focus on particular subject areas that have particular characteristics in terms of knowledge and how that knowledge is used, to build the new realities and assumptions.

The ability to think in the subject will similarly vary across different levels. It would not be expected that students at GCSE would be 'experts' in economics and business but after following a course of study at this level, it can be expected that some capacity to think in the subject will have been developed.

What does thinking in the subject mean? In the context of this thesis, the subjects considered are economics and business and so thinking in the subject involves thinking like an economist or thinking as a business person - approaching problems and new contexts with the skills, knowledge and

thought processes of an economist or business person. In identifying what we mean by 'thinking like an economist' there is an epistemological assumption that such a trait exists not only in economics but in other disciplines. Each subject area and discipline will have its particular blend of skills and attributes that enable the individual to think in the subject. For a mathematician, these attributes may be centred on problem solving and creative approaches to problem solving; to the historian, it may be manifested in the way the individual is able to empathise with people in the past in the context of that era as opposed to imposing modern day attitudes and schema on the past, and an appreciation of the strengths and limitations of different sources of evidence in explaining what has happened in the past. Whatever it is, each discipline will have its own characteristics of thinking in the subject.

Thinking like an economist suggests that there is a way of thinking about economics phenomena that is accepted by practitioners within the subject community. This way of thinking is what separates those within this community from those with an everyday understanding about issues in economics. The approach, methods, conceptual understanding and ways of thinking about the subject help to define what an 'economist' is. Polanyi (1958), suggests that in pursuing a line of study that involves learning and knowledge there will need to be a stage at which individuals achieve equality between themselves in terms of the knowledge possessed by one person and the person who is examining that knowledge.

At one level students might, as emphasised by Marton (1983), have covered certain concepts but have a naïve understanding of them. Their ability to engage in discussion is going to be limited. In short, the ability to 'think like an economist' is dependent on sharing a common understanding of the concepts that form the language of the subject. Some of these concepts will be understood and be used with facility, others might present greater problems because of the abstract, counter-intuitive nature of the concepts.

Polanyi (1958) argues that if we agree with the knowledge that the other

person claims to know and with the grounds on which he (*sic*) relies for this knowledge, the critical examination of this knowledge will become a critical reflection of our own knowledge. In developing this understanding both proponents will be involved in discussing and presenting a 'truth'. That truth, according to Polanyi (1958), is dependent on a common system of collateral facts and values accepted by the subject area.

This argument suggests that there is a way of thinking about economics phenomena that could be described as 'novice' understanding. Rubin (2002) has described these as 'folk economics'; the intuitive economics of untrained persons. Folk economics implies that those taking part in any discussion of an economic issue would not be using the paradigms that the economics community would use in similar discussions. It can be argued that if 'non-members' of this community were engaged in a discussion with 'members' of the community that non-members might have difficulty understanding and following the logic and the approach of the 'experts' because of a lack of exposure to the concepts, methods and approach that characterise the study of economics. The language and concepts that are used within a subject that forms a community of practice is termed 'boundedness'.

Economists recognise in each other the methods, concepts and approaches to problem solving that are specific to the subject (even if they disagree strongly with the outcomes or conclusions reached). Even at the level of study of GCSE there are a large number of concepts that students are expected to understand and use. In the GCSE qualification that is the focus of this thesis there are 96 subject specific terms and concepts in the specification. Many of these concepts are bound in a particular way of operating. Economists use models to help them understand and predict events and issues and arrive at solutions to problems. In this introductory course, students will be learning to think and practise in the manner of scholars of a discipline using a coherently structured body of ideas and procedures to analyse problems as they are defined by that discipline (Davies and Mangan, 2006b).

It is proposed that there is a body of knowledge, a set of skills and methods,

which are used by the discipline to identify, analyse and evaluate problems which constitute 'thinking like an economist'. These problems are 'economic problems' because they are to do with scarcity, choice, decision-making and the outcome of that decision-making. To enable economists to represent those problems and provide commentary and solutions, an understanding of a series of concepts is necessary. Some of these concepts are concrete whilst others are very abstract.

When setting and marking questions for an examination a senior examiner is looking to elicit particular knowledge and skills and enable students to provide evidence of the learning that has taken place. The understanding and skills targeted will include the use of the appropriate models and ways of thinking and practising in the discipline. At GCSE this might not be sophisticated but the early signs of such thinking will be looked for and assessed. The following example might serve to illustrate this, (Ashwin, 2008)

The following two questions appeared on GCSE exam papers in Business and Economics in 2006 and 2007 respectively:

Question 1: To what extent is the development of wind farms the solution to a country's energy problems? (12 marks)

Question 2: Assess the strength of the case for the government reducing the level of taxation to help reduce the level of unemployment. (12 marks)

To be answered well, both these questions require the student to think like an economist. Such a requirement would expect that students consider the costs and benefits involved and then attach some weight to these in order to arrive at a conclusion about decision-making. This approach to economic problems is simple but ultimately at the heart of even the most sophisticated economic analysis and policy-making. In considering the costs and benefits students might be expected to use appropriate concepts and models to help build their analysis. At GCSE the analysis might be relatively simplistic but it will be expected that students show some facility with the approaches that would be expected of someone in this discipline. A number of economics textbooks at the first-year undergraduate level now include a section on 'thinking like an economist' as part of the introductory chapters which emphasises the importance being placed on this skill. McDowell et al (2012) focus on a range of key concepts in Chapter 1 which include scarcity and choice, costs and benefits, rationality, the margin and the distinction between positive and normative economics. Parkin, Powell and Matthews (2008), include a section on 'The Economic Way of Thinking' which includes scarcity and choice, opportunity cost, the margin and incentives as part of Chapter 1. Mankiw, Taylor and Ashwin (2014), include ten principles of economics in Chapter 1 which also cover scarcity and choice, opportunity cost, the margin and incentives and then devote Chapter 2 to 'Thinking like an Economist' where some of the methodology of the economist is covered including the scientific method, empiricism and rationalism, the role of assumptions, models, the distinction between positive and normative economics and judgements in economics.

The University of California (Berkeley), (2011) include a section on 'Learning to think in a discipline' in its Teaching Guide for Graduate Student Instructors. The guide notes that students do not simply acquire facts through a course of study but also disciplinary habits of the mind which enable them to use knowledge. In solving problems, students must be able to control, monitor and self-regulate thinking and this metacognition helps in determining when and how knowledge can be used.

Arnold (2005) provides a detailed account of what thinking like an economist means and entails. His motivation is based on telling stories about how economists think as a method of teaching. This approach, he suggests, engaged students far more than the traditional way in which he had taught economics. The focus of Arnold's book (2005) is on many of the concepts which also feature across other textbook outlines of what thinking like an economist means – trade-offs, opportunity cost, the margin, the costs and benefits of decision making, incentives and unintended effects and equilibrium. Arnold (2005) also includes content on economic methodology, how economists arrive at and test hypotheses and theories to help students understand the mind set of an economist and how they begin to view, interpret and explain the world.

To think in the subject means that the student's thinking has changed and they begin to look at issues from a new viewpoint and with new knowledge. They are able to look at everyday problems and situations and apply concepts and thought processes to explain, hypothesise and draw conclusions. A change in thinking is the distinctive feature of thinking in the subject and highlights the link with learning as defined in this thesis. Students can be assessed at the end of a course of study in their ability to recall facts and use knowledge in certain ways. The assessment of higher order skills can be mechanistic in that students can apply plans and schema to answer questions and gather marks but these skills do not in themselves mean that their thinking has changed and that they now use and apply the knowledge and skills gained in ways that constitute thinking in the subject. The inter-relationships between the aims and objectives of a qualification, the nature of the assessment objectives, and the type of assessment used, influence the way in which the teaching and learning of the subject is framed and the approaches to learning of both students and teachers. These are critical factors in changing pedagogy if it is agreed that thinking in the subject is a key aim. The characteristics of deep learning can be seen as synonymous with thinking in the subject. In economics and business there are key concepts which are essential to acquire in order for the student to be in a position to use these concepts in the methodological framework of the discipline. I will now look at the nature of concepts in general which provides the background for looking at the notion of threshold concepts and troublesome knowledge.

2.4 The Nature of Concepts

Concepts provide an essential means by which human beings communicate. Our knowledge about the world we live in and the way in which we respond to new objects and situations are all influenced by concepts which have been formed by experience. Humans have devised a way of categorising everything we observe and plenty of things we have never observed or experienced. The process of learning and of acquiring knowledge involves building an understanding of concepts and making use of them in a wide variety of ways to enable us to live our lives. In essence, learning does not take place without internalising concepts.

From a relatively early age, children learn new concepts. There is unlikely to have been any formal education as we might understand it to develop this learning. Many children by the age of ten, however, will have a sophisticated understanding of some concepts although they might not be in a position to apply it to more abstract instances. The process of generalisation and discrimination which takes place in concept formation requires that we must be able to distinguish between objects that have different features and respond to those objects accordingly. (Purdy et al, 2001). Learning a concept thus involves learning to respond similarly to related objects even if they differ in some way. Having an understanding of concepts, allows us as humans to cope with different situations in life. It enables us to avoid danger, to respond differently and appropriately to different situations and to communicate effectively with other humans, both physically and remotely, and can result in a changed behaviour and ways of looking at the world.

Concepts in a particular subject in an academic or educational setting are no different in many respects from those used every day. Economics has a wide range of concepts that contribute to the methods and practises that are used to convey information and communicate with others in the discipline. Students arriving at a new discipline have to confront and internalise these concepts; this internalisation can be challenging. There is rarely luxury of time to experience these concepts, to experiment with them and come to know the web of relationships that they have with other concepts in the discipline. Teachers may make assumptions that certain concepts are already understood and this can influence the learning process in different ways.

Some students will find that the difficulties they face in understanding discipline related concepts, and knowing how to use them in the right

context, mean that they never fully come to grips with the subject. Their learning will remain at a surface level where rote memorisation allows them to make some progress. Meyer and Land (2005) refer to this as mimicry. They suggest that mimicry is not just surface learning but rather it represents a state of liminality, a suspended state of partial understanding, or 'stuck place', in which understanding lacks authenticity (Land, Meyer and Baillie, 2010). Liminality implies that students will not be able to transfer their understanding of certain concepts to new contexts, a characteristic of deep learning, and to appreciate how these concepts can be used to help explain other scenarios and issues in the discipline. Their learning will remain largely isolated and behaviour will not have changed.

Deep learning requires the development and fostering of higher order skills and arguments. Underpinning these is a scaffolding of concepts; it is not sufficient to simply understand each individual concept, it is necessary to understand how they interact with one another and how they form part of a relational web. In the two years of a GCSE course there is rarely the time or opportunity for students to build an understanding of fundamental concepts in the way they do with concepts in normal life. The course requires a certain amount of content to be covered and this can be a priority of the teacher rather than focusing on developing conceptual understanding. The relationships between the parts that make up the content that has been prescribed may be less understood and in some cases completely misinterpreted.

In economics and business, students come to the subject with a wide range of prior learning and understanding. Most students embarking on a GCSE qualification at age 14 the will have heard of, or witnessed, a huge amount of economics and business by virtue of taking part in daily purchase and exchange; they simply do not know that they have. The aim of the course is to formalise this prior learning and provide students with the skills, knowledge and concepts to be inducted into the discipline. One of the problems facing teachers is that much of the prior understanding that students have of economics and business concepts is naïve and inaccurate. This prior understanding can interfere with the process of learning how to think in the subject because students have to unlearn these naïve conceptions before re-learning the concepts afresh. The teacher has to help the learner redefine knowledge and encourage the relationship of knowledge with deeper understanding and insight into particular issues and problems. An understanding of the concepts inherent in the subject and the relational structures that bind them together is, therefore, essential.

In non-classroom learning, concepts are likely to have been formed rather than assimilated. Assimilation occurs when we are told about something and expected to understand it; formation occurs where we build the concept into our everyday behaviour more often than not through experience, (Ausubel et al, 1978). The child's exposure to the many instances and web of relationships that help us to make sense of everyday concepts provides the opportunity for internalisation. Many of the concepts that are peculiar to a discipline are not learned in the same way as a child learns everyday concepts; economic and business concepts tend to be the subject to assimilation rather than formalisation. Concepts, therefore, are ways in which we organise thoughts and communicate with others, (Kalat, 2005). They are ways in which we categorise items that allow us the opportunity of building an understanding and communicating that understanding with others.

2.4.1 Theories of Concepts:

The research into concepts is large and complex. It is outside the scope of this thesis to consider this research in detail but the following presents a brief overview of the theories of concepts.

Concept acquisition can be understood as a process in which new complex concepts are created by assembling their definitional constituents. The characteristics that identify a concept can be referred to as their criterial attributes. Categorization can be understood as the process by which a concept is matched to an item to see if its definitional constituents apply to the item. One of the problems associated with this view is that some items are more easily recognised as belonging to a concept than others. There are many instances where it is not possible to specify the definitional attributes of a concept. Various attempts at definitional analysis have failed to arrive at successful and complete categorisations. This suggests that concepts lack a definitional structure and in particular this might be true of the types of abstract concepts found in economics and business.

2.4.1.1 Concepts as mental representations:

Concepts can be seen as an internal system of representation which takes the form of mental images. Purdy et al, (2001) refer to this view as attribute or feature comparison theory. In this theory, concepts are represented mentally in terms of a list of features. Characteristic features are aspects of a concept that most instances share in common. In the case of a market, for example, it must have buyers and sellers and there must be some agreement to exchange using some mutually agreed measure of value usually referred to as price. These features are shared characteristics of all members of the concept 'markets'.

2.4.1.2 Concepts as abilities:

This view states that concepts are specific to so-called 'cognitive agents'those who can think. This ability allows us to distinguish things from each other, for example to understand the difference between a group referred to as 'cats' from animals that are not cats. Such a view implies that whilst cats clearly exist as a reality; non-cognitive agents do not recognise them as 'cats'.

2.4.1.3 Prototype Theory:

In prototype theory the determination of whether something belongs to a category is decided by the extent to which it resembles the prototype of that category, (Rosch, 1978). The idea has its roots in Wittgenstein's idea of family resemblances. Wittgenstein (1968), suggests that many concepts will have certain things in common. This does not mean that one word or concept can be used for each and every case. Instead we can recognise that there are

relationships between the different attributes of a concept. Wittgenstein (1968) uses the example of the word 'games'. There is not one thing, he suggests, that is common to all types of games – Olympic Games, board games, card games and so on - but instead a whole series of similarities and relationships. With prototypes, individuals abstract the common elements of a particular concept and then store an abstracted prototype representation in memory, (Rosch 1978). Prototypes are formed by averaging over large numbers of examples of the concept and then storing an abstracted prototype representation in memory. They become idealised representations of the particular concept. New examples are compared to the prototype to determine whether the new member is an example of the concept.

This idealised representation is something that students coming to a discipline area like economics and business do not have. Many of the 96 concepts included in one awarding body's specification as noted above, will be related to each other, and only when the learner is able to understand these relationships will they be able to think in the subject.

There are some criticisms of prototype theory. Fodor (1998), points out that some categories cannot be described by prototypes. Humans are capable of thinking about imaginary things without ever encountering prototypes of that category. The whole realm of science fiction could be said to be typical of such a view. Kelly (1991), however, points out that many concepts have agreed characteristics and indeed this may well be the case with many of the everyday concrete concepts that most people come into contact with. In economics as in other subject areas, fundamental concepts exist that are abstract in nature; these abstract concepts do not have agreed characteristics; people have to build up their own understanding and thus acquire a personal meaning. Building up a concept involves the use of positive and negative instances - deciding which events and instances belong to the concept and which do not. The problem with abstract concepts is that there may be disagreements about which events can be included and that concrete instances often have to be used before the abstract idea can be grasped.

2.4.1.4 Propositional Theory

Prototype theory is closely related to propositional theory, a theory of knowledge representation (Sternberg, 2003). A proposition is the meaning underlying a particular relationship amongst concepts. This view implies that we do not store mental representations as images. Instead, mental images more closely resemble the abstract form of a proposition. Images are thus secondary phenomena that occur as a result of other cognitive processes. Sternberg (2003) offers the following example to highlight the theory. An image shows a cat sitting under a table. The cat and the table are the objects, the relationship between the cat and the table the subject element. The following propositional representations can be given:

- The table is above the cat
- The cat is below the table
- Above the cat is the table

All the statements indicate the same relationship. If we applied this idea to economics we might make the following statements:

- When price rises, quantity demanded falls.
- Quantity demanded responds negatively to a rise in price.
- Price is inversely related to quantity demanded.
- When price falls quantity demanded rises.

Each of these four statements expresses the same relationship between demand and price. The statements serve to highlight the difficulties that conceptual relationships possess. Such difficulties can obfuscate the path to learning and teachers have to find ways to navigate through these difficulties to help develop deep learning in students.

2.4.1.5 Theory Theory of Concepts

This theory suggests that concepts stand in relation to one another. Morton (1980) proposed that everyday understanding of human psychology resembles

a theory which is then used to predict and explain behaviour in relation to intentions, affect, inherent traits and beliefs. This is a form of categorisation resembling scientific theorising. For example, if a human dresses up in a gorilla costume other humans know that they are not seeing a real gorilla and override the perceptual similarity. Humans have an innate understanding of the essence of being 'human'. In this theory, conceptual change follows the same path as theory change in science. In other words, humans hypothesise, test the hypothesis through observing a number of instances and arrive at a new provisional understandings as a result. Other instances or situations might arise that question currently held theories (concepts) and through the observation and recording process humans arrive at new conceptual understanding, (Weiskopf, 2014).

2.4.1.6 Concepts as Fregean senses

The German mathematician and philosopher, Gottlob Frege, proposed the idea of sense and reference in relation to concepts, (Zalta, 2012). Frege took the example of the planet Venus. The planet can be referred to as the 'morning star' and 'evening star'. However, Venus as the morning star has a different property to the reference to Venus as the evening star. Equally, the example can be given of Marion Morrison and John Wayne. In this case the same referent can have different expressions – John Wayne and Marion Morrison but not necessarily know that they refer to the same person. They present the same person (object) in different ways but have different senses.

Frege claimed that items can have the same reference but different senses, (Zalta, 2012). The sense of an expression is the way it is presented or the cognitive content associated with the expression from which the reference can be picked out. An expression without a referent may still provide meaning to an individual because it still has a sense. The sense is the way in which it refers to an object whereas the reference is the object it refers to.

2.4.2 Concept Formation and Concept Acquisition

Ausubel et al (1978) distinguish between concept formation and concept acquisition. The latter occurs where the criterial attributes of a concept are not inductively discovered but instead presented to the learner as a matter of definition or is implicit in their context. Concept attainment in this case is achieved through assimilation. All too often it may be that the learning that is taking place in classrooms is a form of representational learning, which Ausubel et al (1978) argues follow the concept assimilation process. Such learning may be typical of many students following a course in economics in schools, colleges and universities where the understanding of concepts is incomplete but where students can 'get by' by demonstrating this incomplete understanding through selectivity in the assessment process.

Concept formation on the other hand requires discriminative analysis, abstraction, generalisation and differentiation (Ausubel et al 1978). Such formation may be dependent on the extent to which the concepts concerned are understood by those involved in the pedagogy - the way of thinking in the discipline. The criterial attributes of concepts in economics are important in linking the existing knowledge, understanding and conceptual structures we possess and provide the facility for these new concepts to become meaningful - the concepts have to be internalised.

Ausubel et al (1978) refer to the use of conceptual terms by different members of a given culture and suggest that this does not necessarily mean that there is uniformity in the underlying cognitive content. This may be due in part, they suggest, to developmental immaturity. We see students (who may be regarded as members of the 'culture') using terms like value, opportunity cost and so on but not really having a full understanding of these concepts. Ausubel et al (1978) suggest these represented meanings can be characterised by being vague, diffuse, imprecise or over-inclusive. This can be represented as characteristic of troublesome knowledge (Perkins, 1999).

A lack of conceptual development might be due to a limited exposure to the different manifestations of a concept (possibly due to time constraints

imposed by the need to get through the content of a course) both within the learning environment (classroom etc.) and outside, or the lack of participation in the learning process engendered by too passive an approach to learning. Ausubel et al (1978) point out that many of children's misconceptions come from erroneous or incomplete information or from misinterpretation or uncritical acceptance of what they read or are told.

Few students between the ages of 14 - 16, for example, will have experienced a factory environment and this might have some impact on the ability of a student to fully appreciate economies of scale or the nuances of operations management. The lack of contact with such concrete-empirical experience from which the concepts are derived may be one reason for the lack of understanding (Ausubel et al, 1978). By not having the opportunity to relate the criterial attributes of a concept with these concrete experiences the understanding of many students may remain at an undeveloped and immature level. Ausubel et al, (1978) suggest characteristics of a concept may be restricted to the particular context and experience in which the student first came into contact with it. In examination answers many students are able to express opportunity cost in terms of the choice between an ice cream and going to the cinema (a common example used to introduce the concept) but have difficulties transferring understanding to different contexts and situations and using the concept in much broader discussions characteristic of those who think in the subject. Conceptual development, therefore, might involve a series of reorganisations in which existing concepts are modified as they interact with new perceptions, ideas, processes, emotional experiences, value systems and so on (Ausubel et al, 1978).

Ausubel et al (1978) cite Feldman and Klausmeier (1974) who conclude that if particular attributes of a concept cannot be stated in a way that is understandable to the learner then they should not be included when defining the concept to the learner. This implies that it is necessary to introduce these attributes at a later stage when a fuller understanding has been grasped. Further problems arise when it is difficult to provide any concrete experience of some concepts in economics apart from verbal examples. Abstract concepts such as 'value' can be given verbal explanation and qualities but it is difficult to provide a practical experience of value since in so doing it becomes meaningless. A teacher's understanding of value is likely to be quite different to that of their students and attempts to provide a concrete example might result in further confusion and misunderstanding.

2.4.3 The Nature of Concepts

An understanding of concepts facilitates a representation of reality even if we have a limited view of the exact nature of a specific concept being referred to. Ausubel et al (1978) refer to the difference between the exact experience being described by someone referring to a house and the understanding that the receiver has of that experience. The sender of the information is unlikely to describe in exact detail the particular house being referred to but the mere mention of the word 'house' is sufficient for the receiver to understand the conversation.

If we think of this in terms of how students begin to construct reality in an economics and business lesson when the word 'market' is mentioned, or where the teacher refers to 'elasticity', the representation by the student of that reality might be so different to that being represented by the teacher, that the information being communicated fails to be comprehended effectively.

Entwistle (1983), refers to information processing models and the use of coding procedures in storing information in long-term memory. Information processing, the capacity for coding and analysing incoming perceptions, changes with age. In coping with new information, it is increasingly possible to categorise the unfamiliar in terms of the familiar – to interpret events with reference to previous experience. For many students part of the problem in understanding new information such as that presented in a GCSE economics and business course, is that they do not have this previous experience on which to draw on. Building up a concept involves the use of positive and negative instances – deciding which events and instances belong to the

concept and which do not. The problem with abstract concepts is that there may be disagreements about which events can be included; concrete instances often have to be used before the abstract idea can be grasped.

Pinker (2002), points out that relativist epistemology argues that conceptual categories bear no systematic relation to things in the world but are socially constructed. Examples of such relevant categories are money and tenure. Pinker (2002), points out that in many cases socially constructed categories are based on real truths. Such concepts may differ significantly in their make-up but do have common traits that help those involved in any discussion to enable recognition. The detail of the concepts included in the conversation do not have to be fully understood for the communication to be successful.

The concepts that make up economics and business will have some form of common attributes that allow those involved in the study of the subject to recognise communication. Some of these concepts will be vital for understanding a particular set of behaviours. For example, an understanding of the concept of 'elasticity' is important in understanding why prices for skiing holidays are more expensive during school holidays and why the price of a train ticket is significantly higher between the hours of 7.00am and 9.00am and 4.00pm and 6.00pm than at other times of the day.

2.4.4 Conceptual Understanding

As humans develop a greater understanding of a range of concepts and how they inter-relate with one another, this facilitates more complex understanding. Conceptual understanding is at the heart of learning. As more concepts are acquired, understanding, behaviour and perspectives also change in response to new understanding. Having learnt some concepts it is very difficult to unlearn them, to change and go back to a way of thinking that existed previously. In other words, once a portal is passed, there is a changed reality for the individual.

For students arriving at an introductory course in economics and business as in the case of a GCSE, this will be the start of a process whereby they are inducted into the world of the economist and business and will be expected Page | 95 to learn how an economist thinks and practises in dealing with problems and issues. Davies and Mangan (2006a), refer to 'trajectories of students' learning'. This relates to the experiences that the learner has in the subject which leads to a change of identity.

Learners also arrive with at least some prior knowledge and conceptual understanding. Most students will know of the concept of price or investment, for example, but at an initial stage their conceptual understanding will be naïve. One of the reasons for this is because such concepts will be undifferentiated. If concepts are used in an undifferentiated way, learners will not be in a position to properly analyse issues and problems that are associated with the subject. The challenge is to attempt to get students to unlearn naïve conceptual understandings and move to a conceptual understanding that would be recognised by members of the discipline. This in turn suggests that we are seeking to move learners into a community of practice (Wenger, 2006).

Having tackled some of these initial conceptual misunderstandings, the teacher will have in mind the necessity of building a web of concepts that have relationships with each other. In looking at propositional theory, for example, we can see that any number of propositions can be combined to represent more complex relationships, images or series of words (Sternberg, 2003). These relationships might cover actions (the price rose significantly); an attribute (the level of demand is dependent on factors such as income, taste, prices of other goods and the level and structure of the population); spatial, (demand can be at different levels); or class membership (price is one of the fundamental concepts in economics).

Many concepts can be related to a range of other concepts. For example, the concept of 'elasticity' is not only relevant to economics it has meaning in many other spheres as well. In economics we might relate elasticity to changes in price, income, advertising and the prices of related products but in other disciplines the word has relevance in relation to stress, materials, the properties of a substance and so on. This makes a clear definition of some

concepts very difficult and leads to confusion in the mind of learners. Take the example of the concept 'bald'. Many people would claim to know and understand this concept as relating to hair loss but at what point in the process of hair loss does a person become 'bald'? At what point is a relationship 'elastic'? This implies that there are many examples in life of concepts that are marginal members of a category if we are to view concepts from the prototype perspective, which have 'fuzzy boundaries' (Kalat, 2005).

Given that many concepts have relationships with other concepts, there needs to be ways in which recognition of categories can be organised. Kalat (2005) suggests that it is impossible to think about a concept in isolation; we always have to relate it to something else. This would imply some form of web of concepts linking together related items which help to build understanding. This web of concepts has been modelled by Collins and Loftus (1975) and called networks of associativity. In the model, each node represents a concept and it is connected to other existing nodes through a series of connections through a process called 'spreading activation'. Through such reasoning, concepts can be linked to other concepts in a conceptual network (Kalat, 2005). The conceptual map possessed by an individual helps them to activate or prime the concepts that are linked to it. In the Collins and Loftus (1975) model, the way in which one association triggers another represents an understanding of human cognition. The degree of activation dissipates as it spreads out across the network. Spreading activation occurs faster when two nodes are strongly connected and frequency of use strengthens those connections.

Links between concepts may be made to more specific examples of that concept. In some cases, concepts may be organised in a hierarchy. Davies and Mangan (2005) suggest that the understanding of some concepts in economics may be needed before others can also be understood. The example they use is of comparative advantage. To fully understand this concept, learners have to have an understanding of the concept of opportunity cost. There may, in certain cases, therefore, be some form of hierarchy of concepts in learning. The more distinctive the characteristics of a concept, and the more we understand the category attributes to which we are referring, the more likely we are to recognise something within a category and therefore develop conceptual understanding. The associated cognitive process is important in learning in an academic setting. When a new concept is presented and learned in a classroom we do not have to ensure that this is learned across every category to which it relates. Concepts can be generalised to a category as a whole. It may be necessary to encourage students to use reasoning skills so that they can identify categories and sub-categories. Such a reasoning process enables us to simplify our memory - particularly important when the standard load for short term memory is considered to be around seven items.

Perkins (2006) discusses approaches to concept development in the context of constructivism. Learning in this context involves engaging students in discovery, rediscovery and activity in learning. He refers to this as ritual and inert knowledge. The former is routine and meaningless in character, feeling part of a social or individual ritual rather than having any meaning. It could be argued that an example of ritual learning is where students use concepts without a complete understanding of them. Inert knowledge relates to concepts that we may understand but do not use actively. This sort of knowledge sits in 'the mind's attic' used only when called upon. In this case we might point to the way in which students learn concepts in economics and business but fail to make a connection to the real world they witness and experience every day.

Perkins (1999), links ritual knowledge to the list of misimpressions from everyday experience. Students will learn ritual responses to questions but intuitive beliefs and interpretations surface on problems outside the classroom context. Perkins believes it is necessary to confront students with these misconceptions and the discrepancies in their understanding and suggests that constructivism is one way of doing this.

2.4.5 Conceptual Change

Knowledge acquisition is closely related to the learning of concepts. There is

a debate that places such knowledge acquisition in terms of enrichment or conceptual change. Carey (1991), suggests that concepts are a representation. The existence of innate representations does not preclude subsequent change or replacement. Our representation of concepts change over a period of time and it is possible that one concept can be replaced by another. Enrichment is an explanation of a process of cognitive development that is related to the acquisition of our very early concepts.

This explanation might be relevant to everyday common-sense concepts but when applied to the development of conceptual understanding in disciplines, it might involve more radical conceptual change. The role of the teacher is to seek to replace common sense notions with other, more accurate, interpretations to transform understanding.

For learners, supplanting one mode of conceptual understanding with another is part of the process by which they become inducted into a community of practice (Wenger, 2006). Before that time, learners may well exhibit a different conception than adults who are part of the community of practice. Carey (1991), identifies three different relationships between a learner's conception and that of an adult who we can assume to be part of the community of practice.

- 1. The learner expresses false beliefs represented in terms of the same concept as the adult. In economics and business, this is exemplified in the confusion many learners have of the difference between price and cost. Learners will talk of an item 'costing them £5', which in economics and business terms is an erroneous use of the concept 'cost'. Price is the outcome of a series of relationships and has a close link with the concepts of value and opportunity cost. Cost relates to the outlays of producers in production.
- II. A learner expresses a belief in terms of a different concept from the adult but is definable in the adult vocabulary. The learner, for example, uses the term investment and may well use this interchangeably with the concept 'saving'. For the adult in the

Page | 99

community of practice the use of the term investment means something different although there is a similarity in the underlying meaning of the concept i.e. some expected return for a sacrifice.

III. The learner expresses a belief in terms of a different concept from that of an adult. The learner and the adult's concepts are incommensurable. For example, a learner may have an incomplete understanding of the price paid for an item. The adult in the community of practice has an entirely different perspective based on economic concepts and interpretations such as utility, opportunity cost and value that the learner would not be able to use or comprehend.

Carey (1991), asks how concepts function in a conceptual system, given the contradiction it leads the learner into. Using the concepts of price and cost referred to above as examples of undifferentiated conceptual understanding, if a product 'costs £1' but an individual pays £1.40 for the product, then the learner might be confused about the apparent contradiction. Carey (1991) suggests that part of the answer might be the context in which the learner normally deploys the concept; the more general use of the term 'cost' when strictly the term price is more accurate, does not normally elicit contradictions, for example, when talking to friends or parents, when used in a shop and so on. When faced with questions during assessment that seek the understanding in terms of the differentiation between the concepts the contradiction and confusion can become apparent.

Concepts change in the course of knowledge acquisition (Carey, 1991). When a teacher introduces or uses concepts in economics and business, their understanding of those concepts are different from that of the learner. Carey (1991), argues that it is not just exposure to concepts in the curriculum that makes a difference to conceptual understanding. Many adults may appear to have acquired some concepts from minimal exposure to the topic. There is a suggestion, therefore, that conceptual change is not merely enrichment but is ontologically different. The acquisition of knowledge involves more than changes in belief about concepts. To make this fundamental change in conceptual understanding, learners will be learning to think and practise in the manner of scholars of a discipline using a coherently structured body of ideas and procedures to analyse problems as they are defined by that discipline (Davies and Mangan, 2006b). Davies and Mangan (2006b), highlight three types of conceptual change; basic conceptual change, discipline threshold concepts and procedural or modelling conceptual change. Basic conceptual change refers to the transformation of naïve or 'folk' understanding of concepts to more formal understanding of their distinction and role in economics. Examples of concepts that would fall under this category include the distinction between price and cost, between investment and saving, between wealth and income and between nominal and real data. Such a transformed understanding facilitates a more powerful discipline-based mode of thinking. Such a transformation may well involve troublesome knowledge (Perkins 1999), because concepts as used in the discipline of economics run counter to common-sense notions. Conceptual change in this light can be seen as an acquisition or organisation of schemes of thought associated with the development of disciplinary thought.

Discipline conceptual change relates to the re-working and integration of other disciplinary ideas that the learner has previously acquired. This integration and re-working allows the learner to see interactions and relationships between different concepts in economics. For example, in the area of welfare economics, students will need to integrate their understanding of concepts such as efficiency, cost, value, surplus, shortage, markets, price, opportunity cost and externalities in order to come to a fuller understanding of this discipline concept. Discipline conceptual change provides a unified view of the concepts that integrates them; the whole is more than a simple recognition of each of these parts. Acquiring such a concept transforms the use to which a learner may put their understanding and their perception of the relationship between the concept and other economic ideas.

Modelling or procedural conceptual change are, according to Davies and

Mangan (2005), the enablers without which the learner cannot achieve a deep level of understanding of the discipline concepts. They provide the learner with the ability to understand the type and role of models in economics that enable the user to develop independent analyses or problems. Such modelling procedures might include the idea of comparative statics, the short and longrun, expectations, constrained optimisation, the role of risk aversion and risk seeking in decision making and partial and general equilibrium.

For Davies and Mangan (2005) discipline conceptual change is characterised by being integrative and transformative – two of the key features of threshold concepts. They argue that discipline thresholds are bound together with procedural thresholds. Understanding of the big integrating ideas in a discipline, they argue, cannot be disentangled from the procedures that are used to generate those ideas. We will now look in more detail at the nature of threshold concepts.

2.5 Threshold Concepts and Troublesome Knowledge

The notion of threshold concepts is relatively new and research into the idea has developed apace. The idea was put forward by Meyer and Land (2003) but as the research has intensified the emphasis has shifted from a focus on specific concepts in a discipline to that of learning thresholds in which certain concepts play an important role. Threshold concepts provide a construct to help understand how students travel along the learning journey in a discipline and how they come to develop an 'expert' understanding.

Threshold concepts in all disciplines represent barriers to further progress for the learner (Meyer and Land, 2003). They may be intuitively very difficult to comprehend and in some cases can seem 'intellectually absurd'. Without breaking through this barrier the student will always encounter 'troublesome knowledge' (Perkins, 1999). It could be argued that this troublesome knowledge may well be the source of problems facing students at all levels with questions that seek to assess the understanding of a number of concepts. These concepts will remain fuzzy and hazy in the student's mind throughout the period of their study unless they overcome the threshold concept

Page | 102

concerned, up to which point progression within the subject will remain fragmented. Knowledge can be troublesome because it is tacit, ritualised, inert and conceptually difficult requiring the learner to have to change and use new ways of thinking and new language which they may be reluctant to engage with (Land, Meyer and Baillie, 2010).

Meyer and Land (2003) outline the characteristics of a threshold concept as something distinct within what are usually referred to as core concepts. A core concept is an important concept in the body of knowledge that forms the boundary of the discipline. Understanding of these core concepts allows the individual to become part of the 'culture' that characterises the skills, attitudes, approaches and methods used by experts in the discipline. They will go some way towards providing the means for the individual to think in the subject. Meyer and Land (2003) see them as conceptual building blocks that enable progression of understanding in the subject. A command of core concepts, they suggest, may facilitate some form of understanding but does not necessarily lead to a qualitatively different view of the subject matter or the field.

An understanding of price elasticity of demand for example, would be considered to be a core concept in any study of economics beyond the very basic level. Many students will struggle to come to terms with this concept. To understand elasticity one has to have other building blocks in place demand, supply, income and ratios.

Meyer and Land (2003) identify five important principles of threshold concepts.

- 1. **Transformative** once learned they lead to a significant shift in perception including a shift in values, attitudes and feelings.
- Irreversible once learned the concept would be very difficult to unlearn and may be part of the reason why some 'experts' find difficulty in accepting why students do not understand what might seem, to them, blindingly obvious. Meyer and Land (2003) suggest that whilst irreversibility might be a characteristic it is not necessarily Page | 103

always the case.

- Integrative threshold concepts help provide links with other related concepts helping the learner to see in the round as well as understand particular aspects of a topic. Again however there is a caveat. Acquisition of a threshold concept may increase integrative understanding but may also create more problems in that answers to key questions may not become any clearer.
- 4. Boundedness threshold concepts may be bounded in conceptual spaces that have terminal frontiers at which point new conceptual frontiers will be reached marking the border between different disciplines. This bounded nature is particularly relevant in economics where links with other disciplines is so strong - for example, psychology, politics, sociology, mathematics and history.
- 5. Troublesome threshold concepts can prove troublesome because of the counter-intuitive nature of those concepts and may also be alien or incoherent (Perkins 1999). Meyer and Land (2003) suggest that threshold concepts might invariably have a troublesome element but not necessarily so. Once grasped a threshold concept which incorporates troublesome knowledge may allow the student to move from a common sense understanding to an understanding which may conflict with perceptions that had previously seemed self-evident (Davies, 2003). This seems clearly related to the problems outlined in the integrative nature of threshold concepts in 3. above.

The characteristics of threshold concepts have been criticised as being too vague and not possible to identify empirically (Rowbottom, 2007). Barradell (2013), also notes that identifying threshold concepts has not been easy and that there is a lack of agreement about what the threshold concepts are within disciplines. Rowbottom (2007), argues that statements such as 'probably irreversible', 'possibly often (though not necessarily) bounded' and 'potentially (though not necessarily) troublesome' lacks the precision which allows empirical identification and even whether threshold concepts exist in

any discipline. Rowbottom (2007), continues the argument by stressing that most concepts could be argued to be transformative or at least have the potential to be so. In addition, the transformative nature of a concept will be dependent on the individual and the conceptual scheme that the individual possesses. What is transformative for one individual may not be for another because of the system of concepts that the individual arrives with in the learning experience. A 14-year old student experiencing concepts in economics and business for the first time is likely to have a different transformative experience to that of a first year undergraduate student.

A threshold can be seen as a border or a limit and Rowbottom (2007) suggests that it can be understood that a threshold concept in a discipline is a prerequisite for progression but that if this is the case then threshold concepts begin to look like core concepts rather than being distinct. Davies and Mangan (2005), perhaps set the tone for the way in which research into threshold concepts has developed by suggesting that the ability to construct discipline specific narratives pointed to the way in which students do things in a different way rather than necessarily just how they see things. Rowbottom (2007), picks up on this point and stresses that modelling conceptual change to which Davies and Mangan (2005) refer, need not be transformative in a perceptual sense. The ability of a student to understand modelling as a concept in economics, and how economists frame the world and use models as a means of hypothesising and theorising, is more akin to a threshold as a limit or barrier.

Martin-Hays (2008), develops this idea expanding the definition of threshold concepts to apply more generally to a way of thinking and looking at the world as a result. This is effectively thinking in the subject and Martin-Hays (2008), notes that new ways of thinking allows students to look at new and old concepts in different and new ways. Until the learner crosses this threshold and is able to look at the world differently, their thinking will be 'neglected, resisted or rejected' (Martin-Hays, 2008), which is akin to the idea of troublesome knowledge. Martin-Hays (2008) sees transformation as not just a new way of thinking but a new way of being which will be evidenced in behaviour, manifested in emotions and the way the individual feels. This will be emphasised in the language the individual uses in relation to problems and issues that arise and conversations with other people. Once transformed, Martin-Hays (2008) concludes that 'you cannot be the person you were'.

If specific concepts are to be called threshold concepts, the key is that they are transformative. People within the subject may have a different understanding of the concept of 'value'. One might see value as a representation of the worth of something - the value of increasing spending on education for example. Another may see value as something more bounded - as reflecting a representation of what we are prepared to give up to acquire something, (Wittgenstein, 1968). The implications are that knowledge and understanding will always be ritualistic until threshold concepts have been acquired. This might imply that ritualistic knowledge is inferior and students displaying such knowledge should somehow be penalised in assessment. An understanding of threshold concepts does not, however, suggest that students who have not grasped such concepts will be inferior in some way but they will not have the outlook on the subject and the approach that those who have do - they will not think in the subject. Threshold concepts are not, therefore, once and for all gateways to understanding. Progress through the portal leads the learner to a position where they are able to re-work the understanding that they have previously acquired and incorporated into their thinking; the learners trajectory invokes successively more integration of their thinking, (Davies and Mangan, 2006a). The process of moving through the portal has been described as a state of liminality by Meyer and Land (2005). Liminality, they suggest, implies being altered from one state to another with a subsequent new status and identity within the community. Such a passage can be humbling for the learner. This state of liminality can be entered and there can be times when it will be subject to temporary regression but ultimately there will be no return to the pre-liminal state. The inability on the part of the learner to achieve the new transformed status may lead them to what Page | 106 Meyer and Land (2005) refer to as mimicry or what Ellsworth (1997, p71) calls "stuck places".

Threshold concepts thus can be seen as organising concepts which gives shape and rationale to other ideas being used by a discipline community; they set the boundaries to thinking in a community but can become troublesome for the learner (Davies and Mangan, 2006b). As learners acquire threshold concepts and extend their use of language in relation to those concepts, there occurs a shift in the learner's subjectivity - a repositioning of the self (Meyer and Land, 2005). Threshold concepts, they argue, not only transform thought but also lead to a transfiguration of identity and adoption of an external discourse. This transformation process can be likened to a fluid state bridging the transition between the learner and the portal.

The research into threshold concepts has shifted from a focus on specific concepts in discipline areas, which characterised early research in the field (Mangan et al 2005), to a way of thinking about curricula where specific elements of the discipline represent challenges to the student but which have a transformational impact once they are understood (Land, 2014). Identifying threshold concepts facilitates the planning of teaching and learning and assessment (Flanagan, 2013), and to help teachers reflect on what is being taught, how, why and when to create a focus on critical points in the learning journey which helps to streamline teaching and assessment (Land, 2014, Barradell, 2013).

Thresholds and liminal states imply that assessment may need to be more dynamic rather than a snapshot approach characterised by traditional assessment and can mean students give 'correct' answers but retain fundamental misconceptions. (Land and Meyer, 2010). Land and Meyer (2010), ask how assessment processes can reflect and represent the ontological shift required by threshold concepts. They make the valid point that students can progress to the next stage in their learning (from GCSE to A-level, for example) and find themselves having to negotiate new conceptual complexities whilst still in a liminal state and experiencing troublesome knowledge from the prior course of study. Suggestions made by Land and Meyer (2010) as to what these dynamic modes of assessment might be, include the use of 'write aloud' assessment methods, such as the writing of blogs, learning journals and diaries, and sequential conceptual mappings. These are ways the learner can demonstrate the learning journey they are on and on which the teacher can make judgements about the extent students are progressing on the journey to thinking in the subject.

The research into threshold concepts is on-going and there are developments with other analyses namely signification, schema and capability theory (Land, 2014). Signification reflects some of the issues raised in the discussion of the nature of concepts. Students cannot be assumed by the teacher to arrive at a course of study with the criterial attributes of concepts and a common mental structure. Instead, they are more likely to share words and symbols (signifiers) (Vivian, 2012, cited by Land, 2014). The teacher's role is to lead the learner through a personal cognitive process which may have a range of possible outcomes (Land, 2014) and which may need to be accounted for in assessment. This variation in conceptual understanding within a group has to be seen as being inevitable (Land, 2014) and has to be grappled with by the teacher. It implies that assumptions, and a greater awareness of prior understanding and learning, is an essential part of the teaching process and in the design of assessment.

Schema theory focuses on the idea that knowledge can be organised into segments or schemata and that this stored information represents knowledge about concepts and their associated criterial attributes. The presentation to students of new information such as that in a GCSE course means that the student enters a liminal state which requires new schemas to be developed and that this process represents troublesome knowledge (Walker, 2013).

Capability theory focuses on the ability of students to confront and cope with new and unseen situations and the extent to which the student is able to deal with these situations through a range of experiences in which aspects vary but have commonality (Land, 2014). Baillie et al (2013) have linked capability
theory with the thresholds concept framework to create what they call a Threshold Capability Integrated Theoretical Framework (TCITF). Elements of this approach reflect the idea of thinking in the subject in that rather than the learner mimicking the knowledge and understanding of the 'expert' the learner experiences a personal growth, development and capability which is experienced primarily through variation.

As students' progress through a course such as a GCSE they will all be exposed to threshold concepts to differing degrees. The concept could be transformatory for all but will be a function of the application of these concepts in relation to the paradigm applications. For others the study might not be transformative at all.

2.5.1 Implications of Threshold Concepts and Troublesome Knowledge

Given the discussion on the nature of concept acquisition and threshold concepts, what contribution could an understanding of threshold concepts have for the design of teaching strategies and assessment? The outcome of learning is what has to be assessed so it is necessary to be clear about this outcome. This thesis will assume that deep learning (Marton and Säljö, 1976) and the resultant thinking in the subject is the desired learning outcome and what assessment should be designed to encourage. The characteristics of deep learning have some relevance and overlap with the idea of liminality, thresholds and transformed understanding. It follows that the design of teaching strategies and assessment needs to recognise and encourage interaction with the subject matter to propel the learner on the path to thinking in the subject. This would necessitate ensuring that the content was: accessible,

- that students had some awareness of the subject matter;
- that the assessment was rooted in developing higher order skills of analysis, evaluation and synthesis;
- that the resources were constructed to facilitate a recognition of the parts but yet be able to link together with the whole - relating concepts theories and methods with each other;

- the opportunity to relate existing knowledge with new knowledge and to relate concepts to everyday life where possible, and
- a willingness to relate the materials to other subject areas to help build understanding (Ashwin and Pitts, 2007).

Resources and assessment developed with these principles in mind may contribute to the encouragement of a deep approach to learning and drive the student towards acquiring those threshold concepts that allow them to understand the discipline more completely.

If teaching and learning programmes and assessments are to be designed with the intention of encouraging deep learning and thinking in the subject, one element of this would be to trade-off content for understanding. A typical GCSE specification requires coverage of a large amount of content and it could be argued that the syllabus overload that this presents to both teachers and students leads to surface teaching and learning, a feature highlighted by research from Ramsden (1992), Entwistle (1995), Biggs, (1996), Prosser and Trigwell (1997) and Entwistle (2000).

It could be held, for example, that it is important for a student to explain and understand the core concept of the market in a GCSE economics and business course. Command of this concept involves (at least) an understanding of the elements of demand and supply, the notion of equilibrium, cause and effect, value, opportunity cost, ratios, elasticity and price and cost. Whilst students may be able to offer some explanation of how markets work in examination answers, the range of contexts they are likely to have available and the extent of the variation to which they are exposed is likely to be limited given the extent of the other content which has to be covered. The understanding is likely to remain ritualistic and inert and unable to be applied across a range of contexts. It is posited that the desired learning outcomes assumed in this thesis would be better advanced by reducing the content of the course and focusing instead on a number of core concepts, using variation to enable students to build new signifiers, schemas and capabilities. The assessment thus takes on a different approach focusing not on knowledge, application, analysis and evaluation as the central assessment objectives but on the broader ability to think in the subject.

At the start of this review, learning was defined as a change in thinking as a result of the acquisition of new knowledge, skills and understanding. To achieve a change in thinking, it is necessary for the individual to see the world in new ways, to behave differently and be able to engage in discourse from a perceptually different plane. This definition of learning is reflected in the characteristics discussed about thinking in the subject and transformation. If the student passes through the state of liminality, passes beyond the threshold, then they are able to see the world in new ways, their behaviour changes and they have 'learned'; this in turn links with the ability to think in the subject. Assessment design has a profound effect on the way in which teachers and students approach their studies. Current assessment design such as that for the GCSE business and economics course, is rooted in Bloom's taxonomy and as a result the approach to teaching and learning will be dictated by the characteristics of that taxonomy and the broader desired outcomes of encouraging deep learning and thinking in the subject can be lost. The three research phases of this thesis will look at aspects of the way in which this dominant taxonomy and assessment design influences the learning outcomes in ways that encourage surface or strategic approaches to teaching and learning. This will provide the basis for discussion on ways in which the assessment design can be changed to focus more on the assumed desired learning outcome of deep learning and thinking in the subject.

3. Methodology

3.1 The Research Questions

The purpose of this thesis is to address four key questions which were outlined in the first chapter.

- 1. To what extent do student answers in examinations show evidence of beginning to think in the subject?
- 2. Do teachers have a common understanding of mark schemes developed for assessments?
- 3. Do teachers have conceptions of learning which reflect deep or surface approaches to learning?
- 4. Would changing the assessment objectives in specifications lead to a different focus for teachers and encourage deeper approaches to learning and thinking in the subject?

The research questions inform the research design. Questions 1 and 3 imply a research design which is qualitative in nature. Question 2 involves a focus on assessing questions. Typically, assessment at GCSE, both internally and externally, necessitates the awarding of a numerical mark. The aggregate marks for individual candidates and cohorts are then subject to a mix of judgement and statistical analysis in order to determine the grades candidates are awarded. In judging whether teachers have a common understanding of mark schemes, some analysis of the marks awarded by the participants is required and so Question 2 necessitates some quantitative analysis. Question 4 will be informed by the data analysis and results of Questions 1 to 3.

The inference from the research questions, suggested that a mixed-methods approach would be most appropriate. This mixed methods design focuses on three phases of research each of which required a different approach meaning that the design would include mostly qualitative research but with Page | 112 an element of quantitative in the second phase. The second phase introduced a variable, the mark scheme, which was going to be used to answer the question of whether teacher/markers had a common understanding of mark schemes. In researching the way in which teacher/markers measure student learning, a method was used which sought to reveal whether different mark schemes (the variable) had an effect on the way in which teacher/markers assessed the skills students demonstrated in their answers. There is an implied causal relationship between the variable and the outcome. This research is redolent of that employed by quantitative researchers who use experimental methods and quantitative measures to test hypotheses (Hoepfl, 1997). The outcome of Phase II of the research would include graphs, charts and refer to 'populations' and 'results' (Bogdan and Biklen, 1998).

3.2 Philosophical Assumptions Underlying the Research

The thesis uses a mixed methods approach. The mixed methods methodology arose out of the debate between the positivist tradition and the constructivist movement in the 1970s - 1990s and mixed methods, as a viable research paradigm in its own right, became established from the 1990s. (Denscombe, 2008). Tashakkori and Teddlie (2003, p.x) argue that: "mixed methods research has evolved to the point where it is a separate methodological orientation with its own worldview, vocabulary and techniques". One of the defining characteristics of the mixed methods approach is pragmatism as the underpinning philosophical basis for research (Denscombe, 2008). The thesis presents multiple realities, of students, teachers, markers and of me as a researcher and the claims made to explain the behaviour of these individual perspectives are given through both qualitative and quantitative methods. The qualitative elements of the research are more associated with a constructivist worldview of understanding different realities and the quantitative phase a more reductionist perspective associated with the postpositive tradition (Cresswell, 2015). These two worldviews can coexist through pragmatism or 'what works' (Cresswell, 2015).

Pragmatism in this thesis was a necessary underpinning because quantitative

and qualitative methods alone would not have provided adequate findings for addressing the research questions I had posed (Tashakkori and Cresswell, 2007), Johnson, et al, 2007, Denscombe, 2008). The use of mixed methods is appropriate for social research where application of knowledge and/or human interaction can result in a more complete picture through the combination of different and complementary data sources through the three phases and to facilitate the development of the analysis through building upon initial findings using different and contrasting data and methods (Denscombe, 2008). The primary focus on the first three research questions in this thesis and the consequences of the answers to these questions which informed research question 4 and so, in turn, drove the methods adopted, is a tell-tale feature of pragmatism (Creswell, 2011). Pragmatism implies that the methods adopted are determined by their appropriateness to address the questions (Johnson and Onwuegbuzie, (2004), Mertens, (2015), Patton, 2002).

I approached the research thinking about what I wanted to achieve and had to amend the research method, particularly in Phases II and III, to enable me to achieve the ends. It was, therefore, authentically reflexive and a continual interplay between ends and means and between thoughts and actions (Rayner, 2013). Indeed, as Bouchard, (1976, p402) notes: "The key to good research lies not in choosing the right method but rather asking the right question and picking the most powerful method for answering that particular question". My initial observations on student performance in exams led to research question 1 and action to investigate the question and its consequences. Research questions 2 and 3, were both part of the consideration of the consequences of research question 1. If students were answering questions and securing high marks but not necessarily demonstrating understanding, what role have teachers, the assessment regime (including mark schemes) have in the way questions are answered? Having gathered data on these three areas, I was then in a position to address research question 4 by pulling together the data from the different phases to make inferences. This utilises a version of abductive reasoning within a pragmatic framework, using the results from each phase to construct an Page | 114 answer and series of recommendations flowing from research question 4 (Morgan 2007). Abduction is particularly relevant in this thesis because I have addressed a series of phenomena across different fields (the theory of concepts, approaches to learning, threshold concepts, assessment theory and design, theories of learning and discipline methodology). In collecting data from the first three phases whilst applying concepts from across different fields, I have been able to address research question 4. A pragmatic approach has meant I have not had to impose an abstract theoretical template (deduction) nor just make inferences from data collected (induction) but use the different phases to reason for research question 4 (abduction) (Friedrichs and Kratochwil, 2009).

There are two theoretical bases underlying this research; assessment theory and the theory of threshold concepts. Both attempt to describe a reality but cannot be underpinned by reference to any 'natural law' as would be the case with concepts in the physical sciences. The theoretical bases are rooted in the social sciences. There are elements of psychology, sociology and reference to quantitative methods of analysis. Researchers in the field, as noted in the literature review, have ranged from those in the positivist paradigm (Skinner and Thorndike, for example), to researchers in postpositivist, constructivist and critical influence paradigms.

This research relies mostly on qualitative research. Cresswell and Miller (2000), suggest that part of the way in which qualitative research can establish credibility is through the researcher making clear two perspectives; the lens chosen to validate their studies and the researcher's paradigm assumptions. The lens chosen in this thesis will be clarified in more detail in the section on researcher positionality.

In both Phases I and III of the research, the participants reflect different realities; they will have been subject to influences which shape their view of the world and of teaching, learning and assessment. Their realities will have been socially constructed. The participants in Phase II are also subject to different perspectives and the comments made by some of the participants and quoted in the results, reflect different assumptions and backgrounds and are inextricably linked with their conceptions of learning. Participants in any social act will have different views on the act itself and the outcomes (Shaikh, 2013). Soros (2013), notes that people develop their own beliefs and understanding of phenomena but in forming these perspectives there are bound to be elements of inconsistency and bias. Soros (2013), calls this the *principle of fallibility*. Shaikh (2013), further notes that even the perspective of one individual will be influenced by a wide variety of different values, which in themselves may be inconsistent.

The constructivist paradigm developed out of the study of hermeneutics (interpretative understanding) by German philosophers such as Husserl and Dilthey (Mertens, 2015, citing Eichelberger, 1989). In seeking to interpret and understand human experience (Cohen et al 2001), constructivism places a reliance on qualitative methods and analysis (Mackenzie and Knipe, 2006), with an intention of recognising different views, which Mertens, 2015, p.11) refers to as a "balanced representation of views"; the ontological basis being multiple and socially constructed realities (Mertens, 2015). The epistemological basis is one of a relationship and interaction between the researcher and participants with values and beliefs being made explicit and the findings being created (Mertens, 2015). Phases II and III of this research incorporate multiple realities - those of the respondents - and is also influenced by the position of me as the researcher and the beliefs and values that I bring to the research and interpretation of the results.

The use of quantitative methods in a semi-experimental guise in Phase II of the research makes this one which uses mixed methods (Mertens and McLaughlin, 2003). Mixed methods is a relatively new approach gaining support over the last 25 years (Cresswell, 2015). The mere fact that a researcher gathers data through both qualitative and quantitative methods is not sufficient to make the design mixed methods (Cresswell, 2015). Mixed methods is "An approach to research in the social, behavioural, and health sciences in which the investigator gathers both quantitative (closed-ended) and qualitative (open-ended) data, integrates the two and then draws interpretations based on the combined strengths of both sets of data to understand research problems" (Cresswell, 2015, p2). Mackenzie and Knipe, (2006) note that mixed methods research has come to be associated with what is called a pragmatic paradigm. Pragmatists believe that social science research is not able to uncover real world truth purely by one method, which implies that mixed methods are more appropriate as a research design (Mertens, 2015). Pragmatists recognise that there may be a 'real world' such as that associated with the natural laws of physics but that individuals will have different interpretations of this real world (Mertens, 2015). There is no specific loyalty to any one paradigm and a mixed methods approach can provide insights into the research questions (Mackenzie and Knipe, 2006).

The design used in this research aims to uncover interpretations of assessments of student learning, conceptions of learning, and the way in which thinking in the subject can be advanced and assessed. The research questions are firmly rooted in the belief that there are multiple realities and that I as researcher bring my own beliefs and experience to the study which will influence the way the data is collected and analysed and the recommendations that will be made. A philosophical basis of constructivism and pragmatism both matches the research questions and the chosen design.

3.3 Method and Methodology

The philosophical assumptions outlined above inform the research design. The quantitative element of Phase II developed because the research question aimed to discover whether teacher/markers assess student answers differently if using a different mark scheme. There was a variable involved in this phase of the research and this suggested a quantitative method was appropriate to help answer the question. I had posed a specific question, had a hypothesis to test and hoped to measure the effect of the variable to find an answer. Once the data was collected, I sought to use statistical methods to address the question/hypothesis and use these to interpret the results. Such an approach is consistent with that adopted by a researcher using

quantitative methods (Cresswell, 2012).

Phases I and III were linked to more general questions and the data collection process involved textual interpretation and online interviews. Cresswell (2015) notes that this approach is consistent with qualitative research. Having collected the data in both phases, I had to decide on a thematic analysis using accepted coding and interpretation methods, and to present the findings in narrative form (Cresswell, 2015). The importance of the integration of the different methods is highlighted in Phase II where additional colour and interpretation of the results of the marking was provided by the invitation to respondents to offer comments on the marks they had awarded.

Together the three phases provide a collective whole providing the basis for addressing the fourth research question. Phase I asks whether there is any evidence in exam question answers that candidates are scoring high marks without showing that they clearly understand the subject matter or are showing initial signs of thinking in the subject. The second phase is designed to highlight the hypothesis that a common understanding of what represents a 'good' answer by teacher/markers is open to question and the third phase explores why teachers may focus on learning that encourages planned responses rather than teaching to encourage thinking in the subject and that their different views about student responses might be affected by their conceptions of learning and other external factors. These three phases are designed to establish that there may be some way that thinking in the subject can be encouraged and assessed through the manipulation of the way in which aims and objectives and assessment objectives are designed by awarding bodies and thus provides an answer to research question 4.

Teddlie and Tashakkori (2009) identify two mixed methods research designs, parallel and sequential mixed designs. Parallel mixed designs have the quantitative and qualitative research being carried out together whereas, in a sequential mixed design, the strands occur in chronological order with the questions of one strand (or phase) emerging from or dependent on the previous phase (Teddlie and Tashakkori, 2009). This research is more aligned with the sequential mixed design strand but is not a pure example as described by Teddlie and Tashakkori in that the three phases were identified at the outset as a consequence of the research questions posed and the procedures did not necessarily emerge from the previous phase. They were, however, informed by one another and by a need to adopt a chronological approach due to the fact that I was conducting the research on a part-time basis with the constraints that this presents and so this research can be said follow a sequential mixed design.

The methods used across the three phases were the tools used to collect the data. Decisions had to be made about the most appropriate, effective and efficient way of collecting data to address the research questions. For Phase I the use of examination scripts of students meant that some sort of text analysis method had to be adopted and adapted to suit the data and the research question being addressed.

In Phase II, I also had to adapt the way in which I analysed the data because I wanted to use different mark schemes to identify how consistent teacher/markers were in interpreting student answers to questions. It could be argued that a 'good' answer is a good answer regardless of what mark scheme is being used. This phase of the research presented a number of problems. There is a precedent in analysing and interpreting the efficiency of marking of exam papers (Näsström, 2009, Qingping and Opposs, 2012, Ofqual, 2014c) but in most cases the research has been based on the use of the existing assessment objectives informed by Bloom's Taxonomy (Bloom et al, 1956). In attempting to use two different mark schemes, each based on different ways of viewing a student answer, problems of comparison did arise and had to be addressed as the research progressed.

Phase III of the research utilised the online medium as the means of collecting interview data. Standard descriptions and methods of interviewing had to be modified to address the online mode of collection and did mean that the notion of 'prolonged engagement' with participants (Lincoln and Guba, 1985) is compromised as a method of improving trustworthiness in qualitative

research.

3.4 Position of the Researcher

The use of a mixed methods approach but with qualitative research forming the dominant methodology raises questions regarding validity and reliability. The terms 'validity' and 'reliability' are used extensively in quantitative research but their applicability to qualitative research has been called into question (Lincoln and Guba, 1985, Krefting, 1991, Denzin and Lincoln, 1998, Cresswell and Miller, 2000, Stenbacka, 2001, Golafshani, 2003, Shenton, 2004 and Morrow, 2005).

Researcher positionality is important when discussing the validity and reliability of qualitative research. In Chapter 1, positionality was defined as the social and political landscape occupied by the researcher (Strathclyde, 2014). It was noted that the replication of the methods used in this particular thesis would be likely to lead to another researcher arriving at very different conclusions and recommendations to the ones I will make. Positionality means that the conclusions generated and recommendations proposed are capable of being defended because of the background and perspectives of the researcher.

In quantitative research, validity can be defined as the extent to which the study measures what it set out to measure (Joppe, 2000). Reliability is defined as the extent to which a study would produce the same results if repeated by another researcher (Stenbacka, 2001). For qualitative research, these two concepts do not fit (Krefting, 1991). The knowledge produced by qualitative research is different to that in quantitative research because the paradigm under which the researcher operates is different (Golafshani, 2003) and requires different descriptive terms. Cresswell and Miller (2000), define validity in qualitative research as the extent to which the study accurately portrays the realities of those taking part and the phenomena being described. Lincoln and Guba (1985) provide criteria by which qualitative research can be judged. The terms credibility and transferability are used instead of internal and external validity, generalizability and dependability Page | 120

are used for reliability, and confirmability is used for objectivity (Morrow, 2005). The focus on qualitative research, therefore, is on trustworthiness and authenticity (Cresswell and Miller, 2000).

Guba (1981) put forward a model for assessing trustworthiness which identifies four aspects: truth value, applicability, consistency and neutrality. Truth value is associated with the confidence the researcher has that the findings are true given the research design, context and participants (Krefting, 1991). This can be garnered from the description of the perceptions and experiences of the participants (Krefting, 1991) and in this thesis is particularly relevant to Phase III. Applicability refers to the extent to which the findings of the study can be applied to other similar settings and contexts (Krefting, 1991). The method of interpretation of student answers in Phase I could be applied by other researchers to other examination papers and assessments and the way in which Phase II was set up could equally be applied to other assessments in other discipline areas. Consistency is the extent to which the findings of the study would be replicated if carried out by another researcher using the same methods, approaches and participants (Krefting, 1991). Neutrality is the degree of freedom of bias, motivation and perspectives in the results (Guba, 1981).

Whilst the qualitative elements of this study adhere to these criteria there will be a factor which would mean that some aspects would be different if the study was replicated by another researcher. This difference is due to the interpretation of the results because of the position of the researcher. Cresswell and Miller (2000), note that qualitative research can be validated through the researcher self-disclosing biases, assumptions and beliefs which may impact on the study. Shaikh (2013), suggests that the principle of reflexivity implies that the actions of those involved in the study, including the researcher, will behave in relation to their perspectives and will thus impact on the reality of the study. Soros (2013, p310) clarifies this by stating that "participants' thinking is part of the reality that they have to think about, which makes the relationship circular". Reflexivity describes a two-

way feedback process between participant and observer (Beinhocker, 2013). This study aims to generate some understanding of the assessment of economics and business at GCSE level and of the thoughts and motivations of participants in this system - teachers, examiners and students. Soros (2013) refers to participants' attempt at understanding the world, as the *cognitive function* and the intention to make a difference and advance the participants' interest as the *manipulative function*. These functions can operate at the same time and create interference which means that the independent variable of one function is the dependent variable of the other creating a circular relationship (Soros, 2013). At the heart of this relationship, however, are fallible agents each attempting to understand the actions and perceptions of...(Beinhocker, 2013).

I arrived at this study through my experience and observations on the examination system in England and through teaching economics and business at GCSE and beyond. In seeking to understand why some students seemed to understand and use concepts more quickly and effectively than others, I was seeking a conceptual framework which was provided by threshold concepts and troublesome knowledge. My belief and approach as a teacher and examiner is focused on the intention of encouraging thinking in the subject and of deep learning. In all three phases of the study I am interpreting the data in the light of these beliefs and approaches and so it is inevitable that my interpretations are likely to be different to that of other researchers attempting to use the same methods and following a similar study.

In addition, the way in which I have approached the design and methods used in the study are influenced by personal circumstances. As a part-time researcher I have had to make trade-offs which will affect the way the study is carried out, the data collected and as a consequence, the results. I was fortunate in having access to student examination scripts which had been used for standardisation meetings which I could utilise for Phase I of the research. For Phases II and III, I had to recruit participants and being parttime this meant I had to compromise on the samples that I was able to recruit as well as the number of participants. The sample of markers used for Phase II was limited to 20 in order to be in a position to manage the process of sending out information, securing the participants' engagement with the task and managing the data returned. For Phase III, attempts to gain respondents through an approach which would have secured a more representative sample yielded very little response and as a result I had to adapt and pursue other avenues which meant that the sample I eventually secured may not have been as randomly selected and as representative of teachers involved with teaching and learning at GCSE as I would have liked. Had I been a full-time researcher I would have attempted to secure a larger number of participants for both Phases II and III and probably been able to secure a much more representative sample of both.

The study does, however, represent the views and perceptions of the participants' who *are* involved and is valid in this sense. The study is not aiming to present results which can be generalized across the assessment of economics and business or of other discipline areas. As a piece of constructivist evaluation, the study is my attempt at describing 'what is going on here' (Guba and Lincoln, 2001) and drawing some conclusions about issues which provide the basis for further consideration. These conclusions will be based on the assumptions and beliefs that I bring to the study – that teaching, learning and assessment should be focused on deep learning and encouraging thinking in the subject.

3.5 The Participants

3.5.1 Phase I

Phase I of the study is based on an analysis of student examination answers and utilises content analysis as the method. A key issue in content analysis is appropriate sampling to generate trustworthy data. The 'universe' is the list of all members of the class of documents about which generalisations are to be made (Holsti, 1969; Weber, 1990). In this case the universe is the total number of higher tier examination papers in economics and business for this Page | 123 particular qualification, in a single awarding body for Papers 3 and 4, sat by students from 2002 - 2007. The total number of students in this sample would be approximately 18 000.

The research question helps to narrow down this number to those who gained an A grade in the examination during this period. The total population would be approximately 2 700 (source JCQ¹). The sample studied was taken from student responses whose papers had been viewed by the senior examining team and identified as exemplars for use in standardisation meetings. The aim of the standardisation meeting is for the senior examining team to communicate to examiners their interpretation of the mark scheme and to establish a standard and consistent approach to marking.

Given that the sample will have been discussed and debated by the senior team and used as exemplars they are regarded by the awarding body as being accurate in relation to the marking and the reflection of the relevant mark scheme for the academic year in which the examination was sat. The scripts in the sample come from different years. Monitoring and maintaining standards across different years is a key role of the senior examining team so despite the sample being drawn from different years, there is an assumption that the standard of marking across them can be relied upon as being consistent. In addition, the specification (syllabus) did not change during these years and as a result it would be expected that a student gaining a grade A in 2002, for example, would be of a similar standard to one gaining that same grade in 2007. The number of candidate responses used for this phase of the study was 20.

3.5.2 Phase II

Phase II involves the marking of a selection of student answers by a group of

¹<u>http://www.jcq.org.uk/national_results/gcses/</u> -

http://www.jcq.org.uk/attachments/published/930/JCQ%20GCSE%20Results.pdf Results 2007.

teacher/markers. To recruit markers, adverts were placed on a website for students and teachers of economics and business studies, Biz/ed, (<u>http://www.bized.co.uk</u>), via an email newsletter sent out by Biz/ed and through the annual conference of the Economics, Business and Enterprise Association (EBEA), the professional association for teachers of economics and business. In addition I also placed flyers at in-service training (INSET) meetings I was running for examinations. The result of these efforts was initial interest from 25 teachers.

For those 25 that did respond, I sent an e-mail (see Appendix II) outlining the task, what would be expected of markers, the time scale involved and some information about the incentives available for those taking part. Those that replied saying they would like to be involved were then sent a permissions letter (Appendix III) and asked to read, sign it and return via e-mail. On receipt of the signed permissions letter I then sent the first mark scheme and set of student work to the marker for their attention.

In order to keep track of who had replied and when, who had received which mark scheme and when they had returned the marked work, I created a spreadsheet (see Appendix IV) to log the process. There was a time difference of 36 days between the first person agreeing to take part and the last. Some returned the signed permissions forms the same day (8 of the 25) but others ranged from 1 day (3 people) to 19 days (1 person). There were four other enquiries responding to the adverts but after an e-mail explaining the research task and what was required of markers I received no further communication. Of the sample of 20, 11 were male and 9 female.

3.5.3 Phase III

The sample for this phase was of teachers who were involved with the teaching of GCSE economics and business. I initially placed an advert in a series of subject specific associations and web sites. These, again, included Biz/ed and the EBEA. I also intended to target potential respondents through a random selection of schools in a region of England through accessing contact details via the Ofsted website (<u>http://www.ofsted.gov.uk/</u>) and sending

emails inviting participation in the research. The response to the various means of accessing a suitable population was extremely disappointing. I sent 40 emails to schools selected at random in the East Midlands of England from the Ofsted website but had no response from any of the schools contacted. I received one response from the advert via Biz/ed and 10 via the EBEA conference. I had to seek out other options and at the EBEA annual conference I made contact with a subject officer from Pearson/Edexcel who offered to put the advert into a newsletter to teachers of GCSE economics and business examinations. This route provided some success in getting a further 11 responses. I eventually received a total of 22 completed responses which formed the sample for Phase III. Of the 22 who responded, 16 were female and 6 male.

3.6 Sources of Data

3.6.1.1 Phase I

Phase I of the research focused on content analysis of a sample of student answers to examination questions at GCSE. The sample concentrated on students who were awarded marks at the upper levels of the mark scheme. One of the assumptions made in this part of the research was that the marks awarded to the answers were indicative of the assessment objectives being targeted and that these assessment objectives were targeting not just knowledge but the development of skills in the subject such as application, analysis and evaluation. It can be argued that students who scored high marks on these examination questions might be expected to exhibit the early signs of thinking in the subject given the context of GCSE. Another possible explanation for the reward of high marks for students is that the marking might be open to accusations of being inaccurate or inconsistent. Examiners in national exams are meant to go through training on marking and assessment and are also subject to measures designed to help them understand the standards set by senior examiners (who are responsible for setting the papers and the standards of the examination) and to apply those agreed standards consistently across candidates. It can be argued that if this

process is rigorous, the quality of the marking is not at issue so the conclusion might be that students are being taught or encouraged to construct schemata reflecting the 'test' - in other words, they are able to approach learning in a strategic way rather than an approach to learning characterised by deep learning.

The initial planning for this phase explored ways in which text might be analysed and categorised by looking at some student examination answers and attempting to interpret in some detail what the answers revealed about student understanding of the question and the concepts inherent in the question. I looked at 10 student answer papers to a GCSE economics and business examination and explored the answers in relation to the SOLO taxonomy (Biggs and Collis, 1982). I was seeking to explore levels of and the gualitative nature of the understanding demonstrated. This pilot led to a presentation at the Threshold Concepts Symposium at the University of Strathclyde in the summer of 2006 and the submitted paper was taken forward and published in 2008 (Ashwin, 2008). The categorisation used for the SOLO taxonomy in this paper was generated from a personal interpretation of the different SOLO levels. This proved a useful exercise in developing the analysis of the student responses for Phase I. It made it clear that the way in which I would have to categorise the student answers in Phase I would have to be more robust and subject to cross-checking with external coders.

Content analysis is defined as "any technique for making inferences by objectively and systematically identifying specified characteristics of messages" (Holsti, 1969, p14). The method was adopted in this part of the study to draw inferences about student performance from the data. In this case the data is classed as unobtrusive (Robson, 2002), since the exam papers had been produced for another purpose than for research and are not structured with the needs of the researcher in mind. In this respect it is 'unwitting evidence' in that it allows the researcher to glean information from the data, (Robson, 2002).

The categories had to adhere to basic principles outlined by Holsti (1969)

Weber (1990) and Krippendorff (2013). Categories must:

- reflect the purpose of the research
- be exhaustive
- be mutually exclusive
- be independent
- adhere to the single classification principle.

Identifying appropriate categories and systematically testing them aims to minimise, although never quite eliminate, the possibility that the findings reflect the researcher's subjective dispositions (Holsti, 1969). I identified six categories following a preliminary examination of the data and each was derived from Perkins (1999) rationale for troublesome knowledge. The categories identified are given below.

Categories of Troublesome Knowledge

1. Contradictory knowledge:

2 or more pieces of knowledge are used which are linked by the student but where their use contradicts what would be expected to happen or are linked in an incorrect way.

2. Conceptual misunderstanding:

A particular economic/business concept is used or is referred to which demonstrates misunderstanding of that concept.

3. Partial understanding/knowledge:

A student is using knowledge but the answer suggests that the student has an incomplete or partial grasp of that knowledge.

4. Structured answer to higher tier questions:

A higher tariff question is answered in a way that includes sections covering advantages/benefits/pros, disadvantages/costs/cons, and conclusion which

may have been presented to the student by the teacher as a rote practice.

5. No assessment/evaluation:

A student is required by the question to make some evaluative comment but where points are raised in the answer without any evaluation/assessment given.

6. Unsupported assumptions:

A student makes an assumption of cause and effect without any evidence being provided in support.

Categories 1 and 2 relate to an area of troublesome knowledge associated with conceptually difficult knowledge. The result is a mix of misunderstandings and ritual knowledge often coloured by intuitive knowledge (Perkins, 1999) that students bring with them into the classroom from outside. Students may link two pieces of knowledge together which an expert would not expect to see or which is connected in a way that leads to an incorrect outcome. This will also be evident in straightforward conceptual misunderstanding.

Inert knowledge is exemplified by the learning of concepts which can be used when called upon but where there is little connection made to the real world the student inhabits. This is characterised by category 3, partial understanding, and is also influenced by alien knowledge which comes from perspectives that conflicts with our present understanding. Category 4 is linked to ritual knowledge where students routinely present knowledge without an understanding of the underlying complexity of the material.

In economics and business there is a vocabulary associated with the discipline which students have to come to terms with as part of their learning. This vocabulary helps clarify the mode of thinking and the way in which economists and business people see the world and approach problems inherent in the subject areas. The language and the mode of thinking require students to make and support judgements. Category 5 focuses on a possible inability of students to recognise that the language and way of thinking in the subject requires the ability to present supported judgements which have some balance.

Category 6 is related to tacit knowledge - that which remains personal and implicit (Polanyi, 1958). The existence of tacit knowledge is troublesome in that it arises from the inherent perceived complexity of the subject matter which may seem to the student as being inconsistent. Such knowledge might mean that the student is unable to see subtle distinctions in the knowledge and as a result relies on unsupported assumptions in analysis and evaluation of such knowledge.

Having identified the rationale for troublesome knowledge and established a definition of each category, they had to be operationalised. This was carried out by an initial reading of the scripts and noting down comments on appropriate questions where understanding seemed to be at odds with the apparent level of performance shown by the student on the examination paper. After this process the categories were devised and then revised in the light of the literature on the features of 'good' categories in content analysis (Holsti, 1969, Krippendorff, 2013). This process is known as emergent coding (Haney et al, 1998), which allows categories to be established following some preliminary examination of the data. One of the problems with this method of identifying the categories and subsequent coding is that it involves latent content (Robson, 2002). Latent content is a matter of inference or interpretation by the coder/researcher. Such content and inference can lead to a compromise in trustworthiness from the results of the data. In quantitative studies the issue would relate to content validity and interpreted as the extent to which the scripts are representative of typical grade A students. Phase I is qualitative in nature and so the appropriate terminology to use, as discussed earlier in this chapter, would be credibility or accuracy of representation (Krefting, 1991). Credibility can be established given the assumption that there is monitoring and maintenance of standards over a period of years by senior examiners.

The categories aim to help record instances where students who have secured high marks (and grades) have exhibited troublesome knowledge. Recording the instances of each category throughout the sample of scripts will identify which aspects of troublesome knowledge are most prevalent, if any, and the extent to which able students at this level exhibit troublesome knowledge. The extent of troublesome knowledge can be taken as an indicator that the student is not providing evidence of beginning to think in the subject, which is the research question associated with Phase I.

The relevant code for each category is a simple numerical one, with each category labelled 1 – 6 respectively. Having identified and defined the categories, I then selected two scripts and coded them. Weber (1990), suggests that one test of reliability and validity (and for qualitative research, read 'trustworthiness') is to look at the extent to which the content classification produces the same results when the same text is coded by more than one coder (coder reliability). To test this I produced clean copies of the two scripts and sent them to three colleagues who were examiners along with a sheet outlining the task (see Appendix V), the categories and their definitions. Both scripts were of Paper 4H. Script 1 was student 1, who received 100 marks and Script 2, student 2 who received 80 marks.

3.6.1.2 Phase II

The research question associated with this phase of the thesis was on the common understanding of mark schemes and the extent to which a group of teacher/markers interpreted student answers as evidence of learning. In assessing a piece of work, do teachers look at student answers and recognise evidence of learning on a consistent basis? Do they recognise conceptual understanding of students or troublesome knowledge and is such recognition also consistent? During the course of their work teachers make judgements on student work as part of both formative and summative assessment and the ultimate summative assessment is public examinations. It is not simply in exams where questions can be asked about the efficacy of teachers' conceptions of learning but also in the classroom.

This phase of the research was based on a repeated measures design. Within this design I had different options. One was to ask markers to view students work and use the same mark scheme separated by an appropriate time period. This would measure the efficacy of the mark scheme and provide the basis for comparison of marks awarded over the time period using the same student answers. In this scenario the marks which the teacher/markers had to focus on would have been the same across both time periods and this would have allowed for an obvious comparison to be made. The data set in this research was deliberately kept small scale, partly because of the logistics of organising a larger scale sample but also because it serves to highlight that even amongst a small sample of teachers, there can be sources of disagreement which raise questions about the assumption that teacher/markers hold a common understanding of mark schemes and evidence of learning.

The student answers used in the research are from GCSE public examinations. The marking process for public examinations is regulated in the UK and the awarding of grades based on statistical data as well as senior examiners' judgements. The statistical data used in the exam system is predicated on a standardisation process which purports to ensure that one marker's interpretation of a particular answer is the same as that of others undertaking the same marking. This allows awarding bodies to interpret the marks awarded as continuous data and use parametric statistics to assist in the award of grades.

The selection came from the sample of student examination papers referred to in Phase I of the research. A total of 23 examination papers were reviewed coming from different examination series covering the years 2003 - 2008. It was important to ensure a sufficient number of answers to the same question to provide a range of responses for markers to assess using the different taxonomies. The sample of answers on the papers ranged from short answer questions to those questions targeting higher order skills and requiring some extended writing. I reviewed each paper and made a note of the number of questions available via a tick sheet. The higher tier paper (one designed to target a student entry for grades A^{*} - D) focusing on Perspectives, Paper 4, had three questions of which students are instructed to choose 2. For any sample of these papers there might only be a limited number of questions that could prove to be useful for this stage of the research. For example, there were four examination papers from June 2006 (paper 4H) but following review there were only three of the four students who all answered Questions 1 and 2 and only two who answered Question 3. The value of these papers was, therefore, limited and as a result were excluded from the selection. I aimed to select at least 5 answers to two questions to provide a sufficient range that would enable markers to differentiate between student responses. Paper 4 from June 2007 provided an appropriate selection; there were 5 answers to each of parts 1, 2 and 4 for question 3. Having identified appropriate papers with a sufficient range of answers I then looked at them in more detail to select the actual answers to questions for the marking exercise.

I was looking for a range of answers that would reflect different levels of understanding for both the Bloom and SOLO Taxonomy. Having extremely weak answers would not be of much value for this particular task because they would not give me much of an idea whether the markers were able to identify evidence of learning and differentiate between levels of understanding. My experience of examining is that markers are easily able to appropriately assess weak answers but have more difficulty in agreeing on differentiating understanding and evidence of learning of stronger answers and agreeing a rank order as a result. By selecting answers that were similar in length, used appropriate terminology, and presented some elements of understanding and argument, the task of the teacher/markers to differentiate was made more challenging but more likely to reveal richer data in terms of how mark schemes are interpreted in relation to answers.

3.6.1.2.1 The Exam Questions

I selected two questions, one focusing on the concept of the Boston

Consulting Group (BCG) Matrix (known as the Boston Matrix) in the context of mobile phones and the other on the costs and benefits of London hosting the Olympic Games in 2012. The Boston Matrix is a tool used by business to analyse its product portfolio and aid decision-making. It classifies products in four categories, cash-cows, rising stars, problem children and dogs. The classifications in the matrix are presented along two axes, one showing a growing market and the other a rising market share (the proportion of sales accounted for by the product). A cash cow has a large market share in a mature market, a rising star a low but rising market share in a growing market, a problem child has a low market share in a rising market and a dog has a low market share in a declining market.

One reason for selecting this question was to provide teacher/markers with a relatively short response. It was assumed that this would reduce the opportunity for disagreement over interpretation of the responses in relation to the mark scheme and as a result reduce the degree of variability between the markers.

Question 1 was:

Explain TWO reasons why a business such as Nokia might use the Boston Matrix.

Students answering this question had been given a piece of stimulus material about the mobile phone market and about the Finnish firm, Nokia. The basis of the stimulus material is that the mobile market has become saturated and as a result firms like Nokia have to find ways of generating sales and adding value. It is not expected that students should have a detailed understanding of the mobile phone market but be able to apply their understanding of the Boston Matrix in the context of the market outlined by the stimulus material. The question specifically requires students to give **two** reasons why Nokia might use the Boston Matrix and the command word 'explain' implies that some development is required of the reasons given.

Question 2 was:

Using all the evidence and your knowledge of business and economics, assess the strength of the case for hosting the Olympic Games in London in 2012.

This was one of a series of questions based on the context of London's hosting of the Olympic Games in 2012. The student was given a series of short pieces of stimulus material followed by questions relating to the stimuli. As the stimulus material is presented to the student the questions gradually become more challenging; the question used for this research required students to provide an extended writing answer appearing at the end of the stimulus material. Students would be expected to use their knowledge and understanding of the main concept being assessed to develop an argument and demonstrate higher order skills. They can, and would be expected to, use the full collection of stimulus material given to help develop and support their argument.

Examiners would expect students to present a case both for and against hosting the Games, to use the concept of opportunity cost as the key concept for developing their argument and to arrive at a judgement of whether the case to host the games in London was strong or not as a result of the argument they have presented. Given the time available to students in the examination, detailed answers are not expected but instead examiners would be looking for the way in which the student presents the case for or against the hosting of the Games and how they demonstrate the main assessment objectives being targeted in the context of economics and business understanding.

In summary, the two questions seek to elicit students' knowledge and understanding of the subject matter targeted in the question and to demonstrate a range of different skills, from providing explanation to presenting an argument. In selecting the questions it was important that they provided teacher/markers with sufficient depth of response to apply the mark schemes and thus be challenged to differentiate the evidence of learning and quality of understanding demonstrated by the answers. What might be expected of these questions as evidence of learning? For Question 1, some knowledge of the Boston Matrix would be expected. There may be some assumption that students will have some knowledge of Nokia and the mobile phone market although this is unlikely to have been taught in schools. Question setters often choose a context which they believe will be familiar to students to help them. Students would be expected to apply their knowledge of the Boston Matrix to the context of Nokia (although the phrase 'a business such as...' in the question is a suggestion to students that if they are not familiar with Nokia then they can use a similar business as the context). The command word in the question is '*Explain*' which implies that the student has to offer some analysis to make clear the reasons and to make the issues understandable and intelligible. The use of the plural implies that more than one reason is required.

Question 2 is more demanding. The stimulus material/evidence presented to the student as they progress through the questions take the form of articles, quotes, comments and images which students will be expected to read and use to help address the question. The command word in the question is 'assess' which is a direct instruction to students that this question requires the use of the higher order skills of analysis and evaluation. The remainder of the question implies that there is an argument for and against London hosting the Games and so students will have to present both sides of the argument and then make a judgement about how strong the respective sides of the argument are. The question also implies that students are expected to draw a conclusion to directly address the question.

For teachers, these types of questions are typical of those used to draw out evidence of learning about economics and business. For students who are starting to think in the subject and have begun to grasp key threshold concepts, it can be expected that they will demonstrate evidence of learning and regardless of the mark scheme used, be able to demonstrate the criteria that characterises a high quality answer in the eyes of subject experts - the teacher/markers. For those students who experience troublesome knowledge, answers will demonstrate some confusion in addressing the question. These answers will be of variable quality - where troublesome knowledge is significant the answers will be weaker and where students have worked through some of this troublesome knowledge but not quite grasped threshold concepts, answers might be expected to be stronger.

Teachers could therefore be seen as being arbiters of the extent to which students have begun to think in the subject and grasp threshold concepts through the way they mark student responses. If this assumption was accepted it could be argued that when teachers see evidence of learning, and evidence of learning which is high quality, they will all be able to recognise this and reward it appropriately. It might be expected that there should be a high level of agreement on the quality or otherwise of student answers between teacher/markers especially if they are working to criteria which define the different levels of quality and what evidence is being sought and identified as outlined in the mark scheme.

The SOLO taxonomy has been used more widely across higher education and so teacher/markers may not be as familiar with this mark scheme. In order to ensure that the descriptors in each level of the mark scheme devised for the marking process were as clear as possible but retained the spirit of the way SOLO was developed for use in higher education, I reviewed the literature of those who had used this taxonomy for assessment purposes and noted down the various descriptors that each had used in their interpretation of SOLO. The range of applications of SOLO included examples of student answers in history (Biggs and Collis 1982), student answers on the knowledge and belief about their own learning (Boulton-Lewis, 1995), postgraduate student answers on mental health issues (Chan, Tsui and Chan, 2002), responses from primary and secondary school students in core subjects such as reading, writing and native languages in New Zealand, (Hattie and Brown, 2004), a general introduction to the taxonomy aimed at higher education students (Atherton, 2013) and research with novice programmers (Lister et al, 2006). Having

identified these varied descriptors for each of the SOLO levels I grouped them into the following related areas:

- levels of understanding,
- knowledge,
- development of knowledge,
- analysis,
- development of arguments,
- judgements and conclusions.

From this list I began the process of synthesising the descriptors for higher education assessment and transforming them into a collection of statements in language that would be recognised by teachers working with GCSE students. The descriptors could then be applied to the marking. The main objectives of this process was to create a 'user-friendly' mark scheme which provided a sufficiently detailed range of descriptors to enable markers to appropriately allocate answers to a level within SOLO, balanced by a need for brevity. I aimed to ensure that the SOLO mark scheme was of a comparable length and detail to that of the Bloom mark scheme.

The different statistical tests featured in Phase II were used on the basis of seeking the best fit, if indeed one exists, for such a comparison. I did not want to close the opportunity at the outset by having a design which precluded the possibilities of finding problems and raising questions with applying appropriate statistical tests. This research was designed to be explorative and to raise questions about interpretations of evidence of learning and assumptions about how statistical methods can be applied to comparing and ranking students.

Twenty teachers with experience of teaching GCSE were given the selection of student answers to mark using the two different mark schemes. The sample of markers were sent the student answers by email and post and asked to mark them according to the mark scheme included with the answers. Six weeks later, the markers were given the same student answers to assess but using the different mark scheme. The marks awarded to the student answers were recorded when received at both stages of the process and then analysed using parametric and non-parametric statistical tests. Comparing the results from each marker across the two mark schemes allowed some inferences to be made regarding the extent to which a limited sample of teacher/markers had a common understanding of mark schemes developed for assessments and the extent to which the same marker interpreted the same student answer differently in each case.

3.6.1.3 Phase III

Phase III used online interview questionnaires to gather narratives of teachers' conceptions of learning to reveal some evidence on how teachers conceive of learning in the context of teaching GCSE economics and business and the extent to which there is a belief in deep learning compared to what they actually do in the classroom on a regular basis. Phase III is a non-experimental fixed design. Non-experimental fixed designs define the specification prior to data collection typically informed by a pilot which allows changes to the design, if necessary. The researcher does not intend to change the experience of the participants but to describe and interpret them (Robson, 2002).

The basis for the reason in using an online questionnaire was developed through awareness of the work of researchers as part of the Exploring Online Research Methods project at the University of Leicester (EORM, 2006), and as part of an interpretivist method aimed at understanding, which seeks to study and understand the different ways in which individuals experience, conceptualise, and understand the world around them and their realities, and used in phenomenography (Bowden et al, 1992). Participants are now far more familiar with the Internet and more comfortable with the use of email as a means of communicating and as a result the online environment not only provides for time and space to consider responses to conventional methods of interview such as questionnaires but has the potential to involve a wider range of people who might not otherwise partake in such research (Kralik et al, 2006). The decision to use online interview questionnaires as the method of capturing data in this phase was based on the desire to allow the participants in the research to present a narrative in response to the 12 guestions I developed and thus demonstrate a series of realities which could be interrogated and used as the basis for categorising through thematic narrative analysis, in identifying conceptions of learning. The use of a qualitative data collection method for this phase allowed inductive inferences to be drawn which in turn could be combined with the data collected from the previous two phases to address research question 4 and reflect the pragmatic methodology underpinning the research as a whole. The developments in technology and in particular the Internet, have opened up new avenues of sources of data for researchers which not only yield useful data but also provide for a convenient means of participants being involved which may not be the case with other methods such as personal interviews. I had explored other methods through pilot studies incorporating focus groups at the EBEA Conference and found that there were practical constraints which would have limited the variety and richness of the data I would be able to collect through this method. By utilising a series of similar questions presented online, I gave the participants the time and space to be able to think through their responses to the question without the pressure that exists in face-to-face interviews and in focus groups. This was particularly relevant in some of the questions which are fundamental and challenging, such as Questions 1, 6 and 11. These questions require a chance to think, cogitate and reflect and the use of questions online allowed participants the opportunity of doing so without pressure. This was an important part of accessing the realities of the participants involved.

The development of the 12 questions followed an iterative process based on two pilots. The first pilot involved a series of questions used as the basis of a focus group at an EBEA conference (see Appendix VI) and then a second pilot using questions designed following the experience I gained at the focus group (see Appendix VII). I used feedback from these two pilots, in conjunction with Page | 140 discussions with my supervisors, to decide not only on the final 12 questions to ask but also the order in which the questions would be framed. Question 1, for example, is a challenging and fundamental question in the context of the research and I wanted to ensure that this was at the beginning to focus the minds of participants that this was a set of questions about learning and that I wanted them to focus on their conceptions of learning in relation to all the questions posed. The order of the questions also allowed me the opportunity of checking evidence of contradictions in answers where similar conceptions might be the focus of the question. This is the case, for example, with Questions 1, 7, 9 and 11. A definition of learning, focused on changing thinking, has implications for how learning can be improved and identified and how participants answered these questions would provide some insight into potential contradictory responses.

Evidence from the first two phases might allow some inference to be drawn through evidence in the third phase that teachers revert to teaching schema rather than focusing on deep learning and thinking in the subject, possibly because of perceived external pressures. The use of online questions as the method adopted was chosen following a pilot. I experimented with using a focus-group to address the research question for this phase. A group of volunteers at the annual conference of the EBEA attended a session in which the discussion in response to questions (see Appendix VI) was recorded using a digital recording device and a further volunteer acting as a scribe to monitor the session. Despite the fact that there were a large number of delegates at the conference (around 120) the number of volunteers was small (5 people attended) and the time pressures proved restrictive. The amount of valuable information gathered from the session was limited and the technicalities of recruiting volunteers, getting them to attend on the same day, securing an appropriate venue, paying travel expenses for attendees and the need to ensure that the technology was in place to record the session appropriately would have proved challenging given the part-time nature of my study. This is an example of where the limitations placed on researchers in terms of time and funding means trade-offs have to be made and decisions about the

method of presenting the questions for discussion had to be considered in the light of constraints faced. The pilot helped to shape some of the questions I asked in the online forum and also of the way in which I addressed the analysis of the online interviews. Following the decision to use an online method of conducting Phase III, I undertook a further pilot where I sent 5 participants a series of questions (see Appendix VII) via email and used the questions and responses to further refine the final set of questions and the way in which I addressed the respondents and secured their permission to be part of the research.

The use of email to set a questionnaire and gather research data is relatively new in research given that email itself has only been widely used in the last twenty years but there is literature which provides evidence of advantages and disadvantages. The main advantages are the speed and volume of data collection (Couper et al, 2007, Fleming and Bowden, 2009), cost savings, the flexibility of design, data accuracy, access to research population, anonymity and convenience for the respondent (Madge and O'Connor, 2002). Disadvantages include the potential for sample bias, measurement error, nonresponse bias, length response and dropout rates, technical problems and ethical issues (EORM, 2006). I chose this method because of the speed, cost effectiveness, convenience to the respondent and access to the population. However, sample and non-response bias did prove to be relevant disadvantages. As a part-time student I did have to balance the benefits of using this method against these disadvantages and took the view that this was an appropriate method of data collection.

The data gathered in Phase III is qualitative consisting of a series of narratives told by respondents through the questions. This is situated in the constructivist paradigm used in this thesis in that respondents are expressing their understanding of the reality which they have constructed (Quinlan, 2011); their understanding of what learning is and how they see their role as a teacher. This phase of the research is attempting to describe these realities, to interpret and understand them.

In constructing the questions which formed the basis of the online questionnaire, I need to acknowledge a reflexive process (Quinlan, 2011), whereby my interpretation, understanding and reason will have to be made explicit in the way that I analyse the data, through the themes I identify and the findings and conclusions reached (Quinlan, 2011).

The questions were linked to conceptions of learning identified by Laurillard (2002) with the aim of identifying the extent to which teachers' responses gravitate towards the different conceptions of learning. Laurillard's 'Conversational Framework model' is a transmission model within which the teacher's conception of learning is part of a complex model for learning as shared understanding (Laurillard, 2002). The model outlines the interaction between teacher and student and the 'conversation' which ensues about knowledge and the nature of the subject matter. Teachers use this conversation to modify the perception of learners from which the conceptions of teaching and learning systems can be judged (Laurillard, 2002).

The conceptions are:

- 1. Increase in knowledge/knowledge acquisition/the quantity of information.
- 2. Importance placed on memorizing facts, key terms, using mnemonics.
- 3. Acquiring methods of answering questions or facts to enable questions to be answered and practical application.
- Abstraction of meaning, creating new structures of meaning and knowledge, making sense of information, relating subject matter to other parts of the discipline and to the real world.
- 5. Understanding reality; interpreting, interpretative process to understand reality.
- 6. Changed behaviour, changed as a person, changed thinking and outlook on the world.
- 1 3 are characteristic of surface learning and what you might expect to find

in students who experience troublesome knowledge.

4 - 6 are characteristic of deep learning with 6 being characteristic of those who have passed through the portal to transformed understanding.

The questionnaire (see Appendix VIII) consisted of a total of 12 questions, 6 open questions, each linked to the 6 conceptions of learning outlined above and 6 more closed questions, designed to provide a lead into the discursive questions linked to the conceptions. The responses to the questions were analysed to identify and make inferences about the extent to which teachers reflect the conceptions outlined.

The aim of using open questions via email was to encourage respondents to provide a narrative of their views about learning, their thoughts experiences and beliefs and to explore whether their beliefs reflect the actual approach in the classroom. The analysis of the respondent data comprise a form of narrative analysis. Narrative analysis asks questions related to collective ways of understanding how things should work and can be linked with institutions and 'expert' knowledge.

In reading through the respondent data, the aim was to identify evidence of beliefs through characterising dimensions of the narratives in relation to the defined conceptions of learning. I initially assumed that the narratives would highlight exam boards, school hierarchy, Ofsted, league tables, competitors, exams, government ministers and Ofqual in providing common themes running amongst the narratives - a common feature of qualitative analysis (Robson, 2002). In making these assumptions, I am bringing my own experiences to bear on the research which is one of the reasons for the importance of emphasising reflexivity in social research outlined earlier in this chapter.

Initial analysis of the narratives could identify broad themes related to the six conceptions which could then be collected together to make a more coherent analysis.

The process involved reading through the respondents' replies several times to build familiarity with the language used and to begin to identify the
possible themes that emerge. Having identified a series of themes these needed to be condensed to make them more manageable and the extent to which they link to the 6 conceptions and the division between surface and deep learning. The analysis of the data, therefore, takes the form of thematic narrative analysis (Riessman, 2008).

To test the idea of identifying and grouping themes, a pilot was carried out. Three teachers completed a series of 7 questions (see Appendix VII). Some of the questions were designed to be simple subject based questions such as: "What are the 5 most important concepts that you have to teach students at GCSE (these do not have to be listed in any particular order)"? The pilot study highlighted the need to ensure that the language used in the questions was precise and clear. One of the respondents in the pilot, for example, took the question above to include wider concepts than the question intended. On reflection it was clear that the question could be interpreted in different ways and so was not clear enough in establishing its intention.

As I read through the responses in the pilot study, I noted key phrases and words which either recurred in the respondents' answers or which I judged highlighted an aspect of the 6 conceptions. For example, where respondents referred to a need to 'cover the specification', 'not having enough time to do work that encouraged deep learning' or 'a need to ensure knowledge was absorbed', these echoed the first of the Laurillard conceptions of learning. Qualitative research requires the researcher constructs meaning in the research process through hermeneutics (Quinlan 2011) and in this case I was doing so through interpreting the meaning of the respondents' answers in relation to the 6 conceptions. The questions were designed to aid this process and to tease out the beliefs teachers have about their role, their view of learning and their approach to learning.

3.7 Data Analysis

3.7.1.1 Phase I

The sample included 20 students' examination responses (see Table 3.2). The

sample was taken from a collection of exemplar scripts which included a wide range of abilities. Scripts where students had gained 70 marks or more were extracted and labelled 1 - 20, in no particular order. The reason for this is that the marks gained would have put students in the upper mark range and would have gained them a grade A or above. To achieve grade A, the performance of students should match the grade descriptions that are published by the awarding body. These show the 'level of attainment characteristic of the grade at GCSE' (see Appendix I).

Student No'	Paper	Mark (%)	Year
1	4H	100	2002
2	4H	80	2003
14	3H	80	2003
15	3H	81	2004
16	3H	71	2005
3	4H	75	2005
4	4H	80	2005
17	3H	81	2006
5	4H	73	2006
6	4H	92	2006
7	4H	72	2006
8	4H	70	2006
18	3H	78	2007
9	4H	75	2007
10	4H	92	2007
11	4H	89	2007
12	4H	75	2007

Figure 3.1	Raw marks	for students	in the	sample.
------------	-----------	--------------	--------	---------

13	4H	100	2007
19	4H	83	2007
20	4H	92	2007

The degree of correlation between the coding of these scripts by each coder would give some indication as to the robustness and integrity of the categories as they had been defined. Of the three people identified to help with this part of the research, two initially sent back responses but the third seemed to misunderstand the instructions on the document and merely noted where students had made points and demonstrated certain skills. I returned the scripts and clarified the coding method to the third coder. In the interim period I conducted a simple analysis of the two coders (C1 and C2) who did send back scripts having followed the instructions. I noted the number of times each coder identified one of the categories on a spreadsheet and then inserted my own category identifications. The total number of category identifications was recorded for each script. This allowed me to produce an initial analysis of the degree of agreement between the coders and myself; the initial agreement rate was 66.09%. On receiving the third coder's scripts I added the results to the analysis as highlighted in Figure 3.2.

Student 1	Question	Researcher	C1	C 2	C 3	Agreed
	1ai			3		
	1aii	3,3	3,3	6		4
	1b	6,6	6	6,6		5
	1c	5	3,5	3		2
	1d	3,3,5,6,6,6	3,3,3,5	5	5	7
	3ai					

Figure 3.2 Coder Analysis

	3aii	2,2,6	2,2,6	3	2,3		7
	3aiii	2	1,2				2
	3bi	6					
	3bii	2,2,3,6	2,3,6,6	2,2	3,6		10
	3biii	1,2,3,3,4,6,	1,1,1,2,2,3	1,3,6,	4,6,		16
		6,6,6	,4	6	6		
Total Identifications		29	25	13	8	75	
Student 2	1a	2	2,3	2	2,2		5
	1bi	3	3	3			3
	1bii						
	1ci	6,6	3	6	4		2
	1cii	3,5	3,5	3,6	3		6
	1di	3	3	1	3		3
	1dii						
	1e	4,6	1,4,	6,6	4,4		5
	2ai						
	2aii	2,5	5	2	2		5
	2b						
	2c	3	5	3			2
	2di	3	3	3	2		3
	2dii						
	2e	3,6	1,1,2,3,6	4,6	4		6
Total Identifications		18	17	13	6	54	

			129	93
				72.09

Figure 3.2 shows that the total number of category identifications was 129. The number of instances where the coders agreed with my identification is recorded in the last column. For example, if I had noted category 3 in an answer to a question and the three markers also noted category three for that answer, this was counted as four instances of agreement. The number of times the same categories were identified by all four coders was 93. The total number of agreed identifications was expressed as a percentage. This analysis showed that there was a 72.09% agreement rate between the four coders. This agreement rate suggested that I could have confidence in the coding. I did some further analysis of the extent of the agreement between the coders to assess the extent to which each coder had identified not only the same categories in parts of answers but where the identification was in identical places in student answers. The following is an outline of the question-byquestion instances of coding agreements and differences for Student A. It highlights instances of agreement and difference in the way the categories were interpreted by the two coders who initially sent back correctly coded scripts for each question on this paper.

I believed it important to assess the extent of the agreement between my interpretation of instances of the categories and those of the coders. In any student answer, there might be instances of one or more categories occurring. Different coders might all identify an example of partial knowledge, for example, but the identification might not necessarily be the same. If the categories are well defined and clear then different coders should be able to identify examples collectively. Given the nature of this interpretation of content analysis, there will be some instances where different coders interpret text in different ways. To be effective as a research tool, the categories not only have to be recognised and understood by coders and the researcher but there must be some agreement on the interpretation of these categories in relation to the text being analysed.

The example chosen is a review of the script from Student 1 in the sample and an analysis of where the initial two coders who submitted responses early agreed and/or disagreed with my interpretation.

Student 1, Question 1:

1a i:

Both coder markers noted that the student had commented that a transnational meant multinational. C2 coded this as category 3, the other simply noted it but did not attach a code. I did not code this as I felt it was a statement rather than demonstrating partial knowledge.

1a ii:

Both C1 and I noted the same passages as evidence of partial knowledge (category3) whilst C2 noted a different passage as category 6. The sentence concerned linked communication with diseconomies of scale and could be interpreted as an unsupported assumption.

1b:

All coders noted category 6 in this answer. C2 highlighted the same two instances as I did but C1 only highlighted the second of the two.

1c:

All three coders highlighted the same section as being category 3 and C1 and I identified category 5 as being relevant to this answer which had 'assess' as the key command word in the question.

1d:

C1 and I highlighted two of the same instances where category 3 was given. C2 did not note specific examples of categories in this extended writing answer but instead identified the whole answer as category 5 (which was relevant given the command word of 'assess'). Both C1 and I also noted category 5 for this answer. In addition, I noted category 6 towards the end of the answer which, in one case, C1 identified as category 3. They did this Page | 150 because they had related the paragraph to the key concept in the question, competitive advantage. Whereas I had taken the paragraph as a series of statements which were unsupported assumptions, C1 had linked the conclusion of these statements (that greater productivity and efficiency would lead to an advantage over competitors) as not a clear exposition of the concept of competitive advantage.

Question 3:

3a ii:

C1 and I identified the inaccurate diagram (which had price and quantity on the wrong axis) and also a comment at the end of the answer which suggested that following a fall in demand a business would need to charge more to retain the same level of profit as category 2 - a case of conceptual misunderstanding. In addition we both noted category 6 at the same point towards the end of the answer. C2 looked at the answer as a whole and highlighted the whole answer as evidence of category 3 - partial knowledge.

3a iii:

C1 and I identified the same section as category 2. In addition, C1 had noted a reference to covering losses using internal sources of finance, suggesting retained profit as a source, as evidence of category 1. Using reserves built up from profit in previous years could be a means by which a firm would manage a loss, so I cannot explain why this was categorised in this way.

3b i:

I was the only coder who identified category 6 in this answer.

3b ii:

C2 and I noted the same instance of category 2 whereas C1 noted this same instance as category 3. All three coders noted the same instance of category 2 at the end of the answer. C1 and I also noted category 6 in this same section of the answer.

3b iii:

C2 and I both noted the same two instances of category 6, all three coders noted the same instance of category 1 but my analysis of this extended answer was more detailed in noting instances of category 6 throughout the answer than the other coders. C1 and I both agreed on an example of category 3 but C2 noted category 3 in a different sentence. All three coders noted category 1 in the same place.

In summary, there were a total of 48 instances of category identification and of these 48, 35 instances of agreement occurred either between at least two coders or all three. This represents an agreement rate of 72.9%. Given the overall level of agreement of the three coders and myself over all the scripts of 72.09% the extent of the agreement within this one script gives me the confidence to believe that the categories are robust and the method trustworthy.

3.7.1.2 Phase II

On receiving the marked work, I recorded the results for both mark schemes into a spreadsheet by each of the 20 markers. The teacher/markers were identified by a set of initials and listed in alphabetical order and then allocated the reference 'Assessor 1, 2, 3, etc.' (see Appendix IX). The series of sample observations, (the marks awarded by each of the 20 markers $(x_1,...,x_{20})$ for each student constituted the sample size, n = 20.

As noted above, this phase of the research was a repeated measures design using the same group of participants (the teacher/markers) in all of the treatment conditions (the student answers). In a repeated-measures design, individual differences are not random and unpredictable but can be measured and separated from other sources of error (Gravetter and Wallnau, 2011). By giving the markers the same student answers but with a different mark scheme, separated by a period of six weeks, the intention was to identify variations in the assessment of students' work by the teacher/markers. The null hypothesis is:

 H_0 : - there is no difference in the treatment of student answers using different mark schemes

Any differences among the sample means (an estimate of the population mean) are not caused by the mark scheme but by sampling error or chance.

The alternative hypothesis is:

 H_1 : Markers will assess student answers differently depending on the mark scheme used. On average the use of an alternative mark scheme will have different effects. The mark schemes are responsible for causing mean differences among the samples.

Having represented the data in a spreadsheet, I conducted an analysis based on standard parametric statistical tests. These make assumptions about certain parameters, for example that population distribution is normal and data requires a numerical score for each individual in the sample which can be subjected to some arithmetic manipulation (Gravetter and Wallnau, 2011). Such statistical tests are routinely applied to national examination data and decisions about the award of grades are determined, in part, in relation to the statistical data available to senior examiners and awarding committees. The Bloom-based mark scheme fits the parameter of providing data with a numerical score capable of arithmetic manipulation. The SOLO mark scheme, however, categorises student performance using levels.

Parametric data has to have the characteristic of being continuous; it must be capable of being measured, can take on any value within a range, and one individual's interpretation of each number in the series is the same. In assessing a student answer, there is an assumption that markers can look at the answer and agree on a mark which reflects the evidence of learning and assessment objectives being targeted in the same way that individuals could look at a group of students and agree on the measure of their height. The use of parametric statistical tests in examinations implies that one marker's interpretation of the mark awarded as a result of the assessment is the same as another marker.

There were questions over the extent to which the interpretation of student answers could be analysed using standard statistical tests such as the mean, range and standard deviation if there were wide variations in the marks awarded by markers using the Bloom-based mark scheme. These questions applied equally to the analysis of the marker responses for the SOLO-based mark scheme. Difficulties also arose in attempting to compare the marker responses across the two mark schemes. The mark range using the Bloombased mark scheme for the two questions ranged from 0 to 10 for one and 0 to 16 for the other. In seeking to analyse the results, I was looking at the extent to which there was agreement between markers on student answers within the same mark scheme used. Analysing in-between marks meant looking to see if one marker's interpretation of the value represented by the mark 6 out of 10, for example, would be the same in terms of understanding of evidence of learning as another marker, i.e. does the mark 6/10 means the same thing to both markers?

For the SOLO mark scheme, teacher/markers had to put the answer into one of five levels relating to mark descriptors describing the skills being demonstrated in the answers. Analysing results between the two mark schemes presented difficulties because of the different approaches to awarding a 'mark'. I overcame this challenge by utilising the two sets of data to produce a grid whereby numerical values were converted into a common grade which was then given a common numerical value. For example, using the Bloom mark scheme where the range of marks was from 0 to 10, the range was divided into 6 categories as follows:

0 marks = F

- 1 2 marks = E
- 3 4 marks = D
- 5 6 marks = C
- 7 8 marks = B
- 9 10 marks = A

The equivalent SOLO levels were then categorised as:

No level = F

Level 1 = E Level 2 = D Level 3 = C Level 4 = B Level 5 = A

I used SPSS as the statistical package to analyse the data. SPSS does not allow analysis of alphanumeric data so to overcome this I allocated an arbitrary number to each of the categories to facilitate comparison between the two mark schemes. The alphanumeric terms were allocated a number ranging from 10 - 60 so that all of the marks for both questions and for both mark schemes could now be expressed as a common value. I used a variety of different approaches exploring the comparability both between markers in a mark scheme and between markers across mark schemes, which is described in more detail in Chapter 5.

The process of standardisation in national examinations assumes that the 'live' marks awarded reflect a common understanding and under this assumption, marks are subject to standard statistical tests. In the standardisation process, the mark scheme is explained in detail and examiners have to complete sample papers in discussion with a senior examiner/team leader. Examiners are not allowed to begin marking live papers until the senior examiner has deemed the examiner is marking at the appropriate standard. During the marking process, examiners are subject to checks through different systems. There are weaknesses in any system and the number of re-marks requested by teachers and students which result in changes to grades is testament to a less than perfect system. In 2014, it was reported by Ofqual that the number of enquiries about results had increased by 48% to 443 950 and grade changes as a result of these enquiries had risen by 20% to 43 500, less than 1% of all grades awarded, (Ofqual, 2014d).

These figures, however, do not tell the full picture because it may be that students who get a mark which they, or their teachers do not think is capable

of being challenged such that it results in a change in grade, will not request a re-mark. The extent to which markers disagree on the marks awarded may be more significant than the official figures show but are obfuscated by the systems which awarding bodies have in place to monitor marking.

The practice by awarding bodies of applying statistical tests on exam marks based on the assumption that the data they have is parametric is a feature of the current examination system. If there are differences in the interpretation of student answers highlighted in this research, then the assumption of parametric data cannot be taken for granted. I have included some of the comments which markers made on student answers in instances where there were variations in interpretation to provide an opportunity for some qualitative analysis to support and integrate the quantitative in keeping with the mixed methods design. I also explored some non-parametric tests to overcome the problem of using descriptive statistics to analyse categorical data. These are also described in more detail in Chapter 5.

3.7.1.3 Phase III

The responses to the questions from respondents were in the form of written responses via e-mail. Each response was subject to an initial reading and then three further subsequent readings. All of the responses were read in turn rather than reading one response four times. The reason for this approach was to get a sense of the bigger picture. The first reading gave an initial impression and in doing so I was conscious of how I interpreted the responses to questions. This is where the researcher's personal opinions and experience will influence the way the data is understood which does conflict with Guba's (1981) neutrality aspect in the model of trustworthiness applied to qualitative research. This phase of the research is seeking to discover realities and the human experience in terms of teachers' conceptions of learning and in so doing, is subject to the principle of fallibility as noted earlier in this chapter wherein seeking to understand realities, the perspective developed is going to be subject to biases and inconsistencies (Soros, 2013). Such analysis does not render the research untrustworthy provided the principle of fallibility is borne

in mind when reading my interpretations of the realities that teachers reveal through their answers.

In order to reduce the extent to which personal bias or experience might influence the interpretation, the coding of the responses had to have some rigour and was informed by the pilot study carried out. The pilot revealed that it was possible to highlight a number of sentences or words which reflect a common theme, for example, the following list are words/phrases that were used by respondents in the pilot:

- Absorbing information
- Getting through the specification
- Teaching to the test
- Drumming in answers
- Memorisation
- Preparing students for exams
- Parcelling
- Formulaic
- Tailored answers
- Relevant knowledge
- Repetition
- Practise

These are all words/phrases that can be associated with a conception of learning based on increasing knowledge, knowledge acquisition, information gathering, memorizing, and acquiring methods of answering questions - in other words, a conception of learning associated with surface learning.

As I read through the responses I aimed to replicate this approach by identifying key phrases made by the respondent in relation to the extent to which they link back to the 6 conceptions and the emerging themes. I then grouped these to arrive at the basis for making some judgements about conceptions of learning in the sample group.

3.8 Ethics

The three phases of the research necessitated different ethical considerations. It is pertinent to provide some background explanation to the journey I have taken in pursuing this research. I have been registered at four different universities during this journey as my supervisor has taken up a series of new posts. Having worked closely with my supervisor it made sense to maintain the continuity because he was fully aware of how my thinking had evolved and changed.

The initial guidelines I followed were those of the University of Birmingham where I began my journey. The University of Birmingham's 'code of conduct for researchers' is based on the British Educational Research Association (BERA) Revised Ethical Guidelines for Educational Research (2004). The guidelines were based around 9 elements which are fundamental to the ethical considerations related to research projects.

- 1. Recruitment of participants
- 2. Consent
- 3. Withdrawal
- 4. Confidentiality
- 5. Detrimental effects
- 6. Storage and handling of data
- 7. Harmful or illegal behaviour
- 8. Subterfuge
- 9. Dissemination of findings

Phase I of my research involved the use of examination scripts necessitating consideration of ownership and confidentiality. Phases II and III involved human participants and I had to consider how my interaction with these

people and my responsibilities to them would be informed by the elements above.

Phase I required a submission to an ethics committee. By the time I was involved in the fieldwork for this phase I had moved to the University of Gloucester. This institution also used the BERA guidelines and so I had to discuss with my supervisors whether Phase I of the research needed referring to the University of Gloucestershire's Research and Ethics Sub-Committee (RESC). As Chief Examiner for Economics and Business at GCSE for Edexcel, I had access to a range of examination scripts that had been used for standardisation meetings and which provided a ready supply of data for Phase I. I initially approached Edexcel for permission to use a sample of scripts. In the discussions which followed, Edexcel pointed out that it did not own the intellectual property in the scripts and that it was the candidates themselves who were the owners. It was clear that seeking individual permissions from each candidate used was not an option. The scripts used in standardisation meetings are anonymised before use so it was not possible for me to identify the individuals concerned. Edexcel noted that if I chose to use the data it had no objection but that I had to accept the risk in so doing. In accepting this risk I had to refer back to the 9 elements to ensure that my decisions and use of the data adhered to these principles. I presented the case for using the data to the University of Gloucestershire RESC in May 2008 and the case was considered by the committee on 5th June 2008 (see Appendix X). Approval was received from the RESC and noted in the minutes of that Committee.

For Phases II and III of the research, the participants were all adults. In seeking their participation, I initially sent out permissions forms which detailed the obligations of the participant and my responsibilities as a researcher to them and to the data collected as a result of the participants' involvement in the research. Prior to sending these permissions forms out I discussed the layout and approach with my supervisors to ensure that they adhered to protocol and that I was meeting appropriate ethical guidelines. A copy of these permissions forms are provided as Appendix III.

3.9 Summary

A mixed methods design can also be referred to as a flexible design (Robson, 2002), which is essentially one which generates qualitative data. One of the features of a flexible design is that it allows for a change of approach during the process of data collection and analysis (Robson, 2002). This has been the case with this research, in particular, Phase II, which presented a range of challenges and necessitated a flexible approach as assumptions about the ability to use parametric statistical tests were questioned as the analysis developed.

The importance of researcher positionality and recognition of the principle of fallibility is central to this research. The research questions have been developed as a result of the personal experience I have had as a teacher and examiner and the experience and beliefs I bring to this research not only inform the research questions but will influence the interpretation of the results of each phase. This approach reflects the underlying constructivist conceptual framework of the research. The results will demonstrate multiple realities of the participants in the different phases and the reality, which I will also impose on the discussion of the results, reflects the personal involvement I have in the subject matter under consideration and the way in which the results will be interpreted. Whilst this personal involvement will affect the findings, the methods used to collect and analyse the data have been developed using accepted practices which mean that the research can reflect key principles of trustworthiness as modelled by Guba (1981).

4. Phase I: Results

4.1 Introduction

Before providing detail on the results from Phase I, some further background on the nature of the assessment for the GCSE in Economics and Business is provided.

4.1.1 The Assessment

This section will outline the nature and format of the GCSE assessment, the scripts from which are the basis of the research in this phase. The GCSE examination in economics and business is offered by one of the leading awarding bodies in the UK. Indeed, only one awarding body offers a combined qualification at this level in the two related disciplines. Students in their final year of compulsory schooling at age 16 in England and Wales take a series of GCSEs which form the basis for entry to further education or to the world of work. GCSE economics and business is a qualification that is optional in most schools.

I had been an examiner at GCSE with this particular awarding body since 1994, Chief Examiner since 2003 and Chair of Examiners from 2009 - 2012. My roles involved writing exam papers and the accompanying mark schemes, setting standards, both in relation to the level of the qualification and with standards set in previous years, monitoring a team of examiners to ensure that the standards agreed are met consistently and appropriately, deciding on grade boundaries that form the basis of the award of grades and dealing with enquiries about results and possible re-marks of scripts where such queries are pursued.

The economics and business GCSE up to 2009 consisted of two externally assessed examination papers and a piece of internally assessed coursework. The assessment was divided into two tiers, one for foundation students and one for higher students. The foundation tier papers (1 and 2) were designed

to allow weaker students to access the examination at grades G - C whereas the higher tier papers (3 and 4) enabled students to access grades from D -A*. This research focuses on papers 3 and 4, the higher tier, only. A GCSE examination highlights a range of different ability levels of students sitting the exam and these will be reflected in the grades received following the examination. The grades awarded are associated with a level of performance which reflects the different assessment objectives that form the criteria by which student performance is judged.

Students followed a two-year course of study covering a range of economics and business content. At the end of the course of study students sat two external examinations, each one and three-quarter hours in length. Paper 3 was titled '*Problem Solving*' and Paper 4, '*Perspectives*'. Each contained a variety of questions designed to allow students to demonstrate a range of knowledge and skills based on four assessment objectives derived from the Bloom Taxonomy (Bloom et al, 1956). The assessment objectives for the two papers were defined as follows:

- AO1 Knowledge and understanding
- AO2 Application of knowledge
- AO3 Analysis
- AO4 Evaluation

The various questions that made up both papers were designed to enable students to demonstrate the four assessment objectives. Each assessment objective accounted for 25% of the total marks for each paper. Paper 3 contained a series of pieces of evidence on an economics and business issue. Students were expected to use the context of this evidence in the questions and to use it as support for their answers, where appropriate. The emphasis of this paper was on exploring a problem or problems in Section A, for which there were 70 marks available, and an opportunity to demonstrate some possible solutions to the problems explored in Section B, worth 30 marks.

Paper 4 contained three questions. In each question a series of short pieces of

evidence was followed by questions that built up to provide different perspectives on a topical issue. Students had to choose two from the three questions. Each question was worth 50 marks. Students are expected to recognise and account for different perspectives in relation to the economics and business issue and use this issue to provide a context to help demonstrate their knowledge and understanding.

4.1.2 Mark Schemes

The examination paper was accompanied by a mark scheme acting as a guide to examiners about how to assess candidate answers in relation to the assessment objectives targeted by the question. The mark scheme was drawn up by the senior examiner in charge of the paper initially, and then refined over a period of time as the paper was reviewed and scrutinised. Shortly after students sat the examination, a meeting of the senior examining team was held at which live responses by students were scrutinised. The mark scheme was amended as a result. The amended, final mark scheme was the one which the examining team as a whole would use for the live marking programme.

The mark scheme comprised of mix of a points-based scheme and a levels of response scheme. Points-based mark schemes are used where the degree of judgement required of the examiner is limited and where there is a relatively small number of marks available. For example, questions asking for a key term to be explained using examples would generally be worth four marks; there would be two marks for the definition and a further two for the example used. Examiners were given exemplar work to help them differentiate a one mark definition from a two mark one and where an example was worth two marks rather than one, and so on. For questions which carried a higher tariff and where expert examiner judgement was required, a levels of response mark scheme was used. Here, student responses were placed into a level in relation to the extent to which they met the assessment objectives being targeted by the question. A 12 mark question, for example, was likely to be heavily weighted towards the higher order skills of analysis and evaluation. To reach the upper levels, which may

be level 3 (7 - 9 marks) or level 4 (10 - 12 marks) the mark scheme had to specify what sort of evidence of analysis and evaluation skills the student had to demonstrate to reach those levels.

4.2 The Results

Following the coding of each script, I recorded the instances of each category noted on a spreadsheet (see Appendix XI). There were a total of 275 questions answered by students in the sample. Each question was linked to the number of categories identified (if any). The number of each category instance was then noted against the category number. For example, if a question had two recorded instances of category 3 then this was recorded as '2' under the category 3 column. The total number of instances recorded were then summed and a horizontal and vertical cross check of the corresponding totals taken to ensure that the calculations had been correct. There were a total of 267 category instances recorded. The spread between each category is outlined in Table 4.1.

Table 4.1 Breakdown of Category Instances in Relation to Total Instances.

Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Total
19	46	61	24	24	93	267

Representing these totals as a proportion of the total instances gives the percentage breakdown indicated in Table 4.2 and shown as a horizontal bar chart.

Table 1 2 Scri	nt Analysis.	Catonorios as	a Proportion	of Total Instances
	pi Analysis.	categories as		UI TUTAI IIISTAILES.

Category	Percentage (%)
1	7.1
2	17.2
3	22.8
4	9.0
5	9.0
6	34.8
	100.0



It is clear that category 6, unsupported assumptions, was the most common instance recorded. Every student apart from one (Student 10) in the sample exhibited this category at least once in their answers with 12 students having Page | 165

this category recorded at least 4 times across the questions answered. Category 3, partial knowledge, was the second most common with 17 out of the 20 students in the sample recording at least one instance of this category. Conceptual misunderstanding (Category 2) was also relatively common accounting for 17.2% of the total instances identified.

Examples of some of these categories follow:

An introductory question on the Higher Tier paper in 2006 asked students to: 'Explain the term 'import''. Student 8's answer showed some evidence of partial understanding (category 3) in commenting: "Import is when a product is brought into a country or the EU from another country". This is partial understanding because there is no recognition of the flow of money between the two trading partners - an import is the purchase of a good or service from another country representing an outflow of funds for the UK.

Later in the paper, this student demonstrates further misunderstanding of the concept of imports in response to a question on why the EU might impose a tax on imported goods from other parts of the world. Student 8's response suggested that taxes were imposed to prevent importers from "...fixing the price of their goods to prevent unfair competition".

Student 2 showed an example of conceptual misunderstanding (Category 2) as part of an answer which addressed factors that could contribute to the success of launching a new product where the context given was the launch of Microsoft's Windows 95[™] operating system. Student 2 linked Microsoft's success with Windows with a market niche. The definition of a market niche is a small part of an overall market. Given that Microsoft dominates the market for operating systems this was a clear example of not only a conceptual misunderstanding of the concept of market niche but also partial understanding of the context supplied.

Student 1 showed an example of partial understanding and of contradictory knowledge in offering a definition of diseconomies of scale which: *"arise when expansion leads to inefficiency and lower productivity"*. To demonstrate complete knowledge and understanding the student needed to Page | 166

refer to the impact on unit cost; that unit costs rise when diseconomies of scale set in. Productivity is output per factor of production per time period and can fall for many other reasons than diseconomies of scale. Student 1 also demonstrated partial understanding of the concept of competitive advantage by noting that: *"The firm with the best mix of the 4Ps* [of the marketing mix] *will have a competitive advantage over other firms"*. The essential features of competitive advantage is that it has to be distinctive and defensible - all firms will have some elements of the marketing mix (price, product, promotion and place) but simply having the right combination is not sufficient to confer competitive advantage. If other firms in the market are able to easily copy and action the elements of the market mix of other firms then competitive advantage does not exist.

Student 1 gave another example of partial understanding when explaining how a decline in the market for children's building bricks might affect price and quantity sold. In their answer, Student 1 noted that a reduction in demand would lead to a lower market clearing price (correct) which in turn meant lower profits for the firm (not necessarily) and then concluded that the firm needs to "...charge more to retain the same level of profits but this could cause even less demand". There is evidence of confusion in the answer; the question makes it clear that the market for children's bricks is falling and the requirement of the student is to show how this change in market conditions would affect price and quantity sold. The student recognises that the fall in demand would lead to a fall in market price but does not comment on the quantity sold. Instead the student makes a specific link with a fall in price and the effect on profit. The question asks that students '... use supply and demand analysis...' to address the answer which implies that the focus of the assessment is on the ability of students to understand basic market models. The amount of profit a firm generates is not shown by the market model of supply and demand. The student suggests that the firm will need to increase price in order to maintain profit which appears to ignore the context of the question that the market is in decline and demand is falling. Increasing price would not, in itself, lead to a rise in profits. The student does attempt Page | 167 to qualify their statement by noting that raising price might reduce demand further but does not offer any further development of this point. Overall, the student shows some confusion over the model being used as the basis for the question and some misunderstanding of the nature of the context.

As noted, the most common category was Category 6, unsupported assumptions. Student 8 noted in response to a question on the regulation of monopolies that: "*Monopolies face limited competition…*" - dependent on the market this may or may not be accurate but there was no attempt by the student to qualify their assumption by reference to a context or an example. The same student made further unsupported assumptions later in the paper in response to a question about the benefits to a less developed country of investment by a multi-national company. The initial part of the answer given by Student 1 made two unsupported assumptions: "*The investment by multinationals will create more jobs for the local people*", (not necessarily the case), followed by: "*This means more people will have a larger income that (sic) what they had before investment…*". Given the fact that the first assumption made was unsupported and its accuracy questioned on a number of fronts, the linked second assumption then becomes even weaker.

Student 1 in answering a question on investment in new technology by Tata Tea, suggested that "...new technology would help Tata Tea to reduce costs... (a major assumption) as better technology in machines would mean less staff required..." (a further unsupported assumption). There was little evidence provided that the student understood that investment in new technology involves considerable cost in the short-run and that new technology does not necessarily mean that a firm will shed labour, especially in a market such as tea production which relies heavily on labour in certain parts of the production process. The same student gave a further example of an unsupported assumption in response to a later question on the paper about how the success of retail outlets during the Christmas sales period of December 2000 would affect the economy as a whole. Student 1 noted that there would be "...more wealth generated, increasing bank investment, which would too benefit the economy". The student appears to confuse the terms 'income' and 'wealth' (income being a flow and wealth a stock) and how this relates to retail outlets experiencing higher sales over the Christmas period and how higher retail sales lead to an increase in bank investment. The student followed up with "It would help to stabilise the economy and even strengthen it" but gave no indication of how higher retail sales could do either of these things.

4.3 Reflection on Coder Analysis: The Difference Between Content Analysis and Text Analysis

The differences between the text analysis in this research and content analysis as defined in Chapter 3 is that content analysis developed as a means of quantitatively measuring specific words or sentences in text, television programmes, movies or speeches, particularly in the realm of politics. The researcher may be looking for instances where particular words or phrases occur and be able to quantify the instances to make inferences (Stemler 2001). Such inferences might signify a likelihood of a particular behaviour, set of beliefs, intention to make a point and so on.

In such cases of content analysis the definition of categories is very clear noting instances of the word 'republican', 'democracy', 'inflation', 'leadership' etc. is something that can be carried out manually or with the help of computer packages and is relatively easy to quantify.

In this part of the study, content analysis is being used to analyse content from the point of view of identifying instances of misunderstanding, misconception, partial knowledge and so on. These instances occur in sentences and were not the main intention of the student who wrote the words. In fact, it could be argued that the student intended the exact opposite of what I was looking for. Sentences can be categorised in the way defined above but sentences are capable of being interpreted in different ways. In addition, different individuals may spend different amounts of time focusing on the task in hand. As a researcher my focus on the task had to be more robust and rigorous than that of colleagues who agreed to act as coders. Page | 169 The fact that one of the coders completely ignored the instruction sheet may lend some weight to such a view. The motives of the coders and the focus they have on the task may, therefore lead to widely differing interpretations of the categories.

4.4 Summary and Conclusion:

This first stage of the research has discovered that students who are considered able at this level of study demonstrate a range of problems consistent with troublesome knowledge. This is exemplified by the prevalence of unsupported assumptions. For example, Student A noted that "Communication is usually the main contributor to diseconomies of scale. [A firm] can overcome this by regular meetings to discuss ideas"; Student D commented in relation to reasons why monopolies might be subject to regulation: "Monopolies can set the price at which goods are sold because they have the power in the market and customers will have to pay the price set, even if the price set is extremely high and unfair". The analysis of this particular sample suggests that students getting high grades at GCSE can experience troublesome knowledge in articulating answers in examination papers and do not show early signs of thinking in the subject and the beginnings of a transformed understanding, even though they scored high marks and achieved at least a grade A in the examination. This may suggest that students are exhibiting the characteristics of surface or strategic rather than deep learning. Student D demonstrated how marks can be gained through providing a schematic answer which gives pros/cons and a conclusion. The answers given to questions 1g and 3g, the high tariff questions, both followed this particular approach to answering the question. By offering some balance and a conclusion, the answers given by Student D would at the very least meet the criteria for a Level 2 answer in the levels of response mark scheme.

Such an approach to learning may have some roots in the learning experience of students in the classroom and to the assessment regime they are subjected to. This led to the focus of the second research question and an attempt to find out if there are any differences to the way in which student learning is interpreted with different assessment regimes.

The next stage of the research might provide some indication of whether an assessment structure based on interpretation of Bloom is the best way to assess learning and understanding if it is assumed that the desired outcome is deep learning and thinking in the subject. An alternative taxonomy may produce very different and preferred outcomes in teacher's conceptualizing and assessing of the evidence of learning and understanding. This will provide a platform to find out whether there are any patterns associated with student experience in the classroom and the assessment regime used, and the third stage of the research, investigating teachers' conceptions of learning.

5. Phase II: Results

5.1 Introduction

In the first phase of the research, a selection of student answers to examination questions was subject to text analysis. There was some evidence that students were able to access marks based on schematic answers rather than demonstrating the deep understanding that might be expected if the student had grasped key threshold concepts and was beginning to think in the subject. In the second phase of the research I wanted to look at how teacher/markers interpreted student answers using different mark schemes. At the upper levels of these mark schemes, the sort of skills required of students would be typical of those who could be expected to be thinking in the subject and have begun to acquire threshold concepts whilst those at lower levels would be expected to exhibit troublesome knowledge. If students are clearly demonstrating evidence of learning in the form of a grasp of threshold concepts and thinking in the subject, then it might be expected that teacher/markers could recognise this evidence of learning and that there would be some collective agreement demonstrated.

In this phase of the research a basic question in relation to threshold concepts and troublesome knowledge is being asked. Teachers of economics and business have some expertise in their subject area and are also trained in assessing evidence of learning. In many instances, the evidence of learning is a written response to a question set by a teacher or in the case of exams, by an awarding body (which in practice is likely to be written and have the standards set by a teacher acting as a senior examiner). The written response contains evidence of the student's understanding of the subject matter and the particular question asked.

Marking students' work is part of teachers' daily activity. In examinations, learning outcomes are defined by the assessment regime being used. In GCSE examinations in economics and business, the regime is based on Bloom's Taxonomy (Bloom, et al, 1956). Teachers will seek to develop a familiarity with the taxonomy and the assessment criteria to use in the classroom to reflect that which students will ultimately be tested against. It can be argued that teacher/marker familiarity with the taxonomy should mean standards across the assessment are valid and reliable, that is the assessment measures what it sets out to measure (evidence of students' learning) and a similar outcome would be secured if the student response was marked by different people at different times. Analysis of this element of the research will give some indication as to the extent of this validity and reliability.

Characteristics of deep learning appear in the 'aims and learning outcomes' of the GCSE Subject Criteria for Business Subjects (Ofqual 2011b). These characteristics include a requirement that GCSE Specifications should:

- actively engage in the study of business and economics to develop as effective and independent learners, and as critical and reflective thinkers with enquiring minds;
- use an enquiring, critical approach to distinguish facts and opinions, to build arguments and make informed judgements;

Ofsted (2011, p13) notes that a particular feature of outstanding teaching and learning is evidenced where:

"a strong focus on developing students' understanding of key concepts and use of technical vocabulary; their ability to present arguments (both orally and in writing); their ability to apply their knowledge and understanding to unfamiliar contexts and to analyse and evaluate was present", and where there was evidence that teachers were: "asking probing questions that aimed to confirm understanding and extend and deepen thinking".

Both Bloom and SOLO are designed to measure the extent of student learning. There is a difference in the approach used by each taxonomy. Assessment criteria based on Bloom's Taxonomy tends to separate out knowledge, application, analysis and evaluation with the result that teachers might look at a student answer in an atomised way, picking out elements of each and awarding marks on identification. SOLO, on the other hand, places an emphasis on the quality of the learning displayed which implies markers should view the answer holistically. This can lead to a difference in the identification of evidence of learning in the marker. By presenting the same pieces of work to markers and asking them to use the two different taxonomies, my aim is to find out if there is a difference in markers' perceptions of evidence of student learning and to ascertain whether, and in what ways, each taxonomy affects markers' perceptions.

5.2 Mark Schemes

Since the mid to late 1990s awarding bodies in England and Wales have published examination mark schemes. Teachers use these published mark schemes in various ways. In particular, they give a clue about how examiners are thinking and are used to inform teachers' planning of lessons and learning. Some teachers will use them with students as part of peer assessment or to reinforce learning objectives. Mark schemes also provide a way in which understanding of evidence of learning can be standardised giving the assessment validity and reliability.

In its Code of Practice, Ofqual (2011d, p.14) notes:

"The standardisation process is designed to make sure that all examiners mark candidates' work consistently and accurately. It establishes a common standard of marking that should be used to maintain the quality of marking during the marking period".

It is the responsibility of the awarding organisation to:

"... ensure that all examiners have a well-founded and common understanding of the requirements of the mark scheme (appropriate to their responsibilities) and can apply them reliably". (Ofqual 2011d, p25).

In a mini-pilot designed to identify the difficulties of getting agreement on the breakdown of assessment objectives in examination questions, I presented an examination question from an advanced level Business Studies examination to a group of teachers and examiners at the EBEA conference in 2006. I asked the delegates to decide how many marks out of 14 would be allocated to knowledge, application, analysis and evaluation for the following question:

Although the long-term potential for air travel seems good, economic and other problems can hit demand in the short term. Discuss suitable strategies which Airbus needs to have in place for coping with unexpected variations in demand.

(Source: AQA Business Studies Unit 6 paper, June 2004).

One of the delegates was the examiner who actually set the paper and thus had devised the mark scheme and decided on the allocation of marks across the assessment objectives. There was little agreement around the room on the proportion of marks appropriate to each assessment objective in comparison to the actual published split by the awarding body and the examiner who set the paper also offered a different split than the one published which was as follows:

- AO1 2
- AO2 2
- AO3 4
- AO4 6

The lack of agreement suggests that the interpretation of the assessment objectives targeted by a particular question is open to a considerable amount of subjectivity.

Mark schemes based on Bloom generally comprise a points-based mark scheme typically (but not exclusively) used for questions where the number of marks available is below 6, and a levels-of-response mark scheme where the number of marks available is higher than 6. A levels-of-response mark scheme provides a series of mark ranges (levels) relating to the demonstration of higher order skills. Simple analysis and evaluation may be characteristic of a level one answer and the more developed these skills are the higher the level and thus the marks which students can access. The SOLO mark scheme is a series of 5 levels which require the marker to look holistically at the answer and focus on the quality of the learning displayed. At the higher levels the quality of learning demonstrated is explained in terms which are very similar to the characteristics of deep learning.

A key difference in the two mark schemes, therefore, is that markers might look at the Bloom-based mark schemes, which is the one familiar to most teachers at this level, and interpret student answers differently to the SOLO mark scheme which focusses on the quality of the student answer. A mark scheme focusing on the quality of student understanding is more likely to be one which would be used to assess the extent of thinking in the subject and grasp of threshold concepts, and reveal the existence of troublesome knowledge rather than a mark scheme which focuses the marker's attention on identifying particular traits in the answer which would be typical of a points-based mark scheme.

Communicating this difference to markers prior to the initial marking of work was deemed important. In the documentation sent to markers (see Appendices XII and XIII) a guide to using both mark schemes and the underpinning philosophy was given. The explanation of the SOLO mark scheme was particularly important because teacher/markers may not have encountered or be familiar with the SOLO taxonomy, primarily because SOLO has been used almost exclusively in higher education.

In the explanation I included a paragraph pointing out that the assessment has to be seen in the context of the average 16-year old taking an examination with the inherent pressures that this brings. I included this proviso to remind markers to use the mark scheme rather than interpreting answers in their own way and thus be excessively harsh in judging the evidence of learning. The markers were given the questions, the evidence that candidates would have seen on the exam paper and 8 answers for Question 1 and 5 answers for Question 2.

5.2.1 The Bloom Mark Scheme

The mark scheme begins with an introductory paragraph outlining to markers Page | 176 the nature of the mark scheme (see Appendix XII). It was decided to use the revised assessment objectives for GCSE economics and business which merge analysis and evaluation into one, which was the format in use at the time of the collection of this phase of the research. The three assessment objectives were: Knowledge and understanding (AO1), application of knowledge (AO2) and analysis and evaluation (AO3). Teacher/markers were also given the number of marks that could be awarded to student answers for the two questions. For question 1 on the Boston Matrix, the marks available were up to a maximum of 10 and a points-based mark scheme was provided. This particular questions was structured such that it could have been two, five mark questions, and so a points-based mark scheme was appropriate. For Question 2 on the Olympic Games, 16 marks were available using a levels-of-response based scheme.

Using the points-based mark scheme for Question 1, the marker is looking for evidence that the student can identify two reasons for using the Boston Matrix in the first instance. This is a demonstration of knowledge (AO1). The command word 'Explain' implies that the student needs to offer some analysis to develop the reason/s identified (AO3) in the context of Nokia (AO2). There are 2 marks awarded for the demonstration of knowledge (the two reasons given), 2 marks for the appropriate use of context and the remaining 6 marks are for the demonstration of AO3, analysis (3 marks for the development offered for each reason identified). Structuring the mark scheme in this way ought to mean that teacher/markers will be able to recognise appropriate reasons and that the answer uses the context of Nokia. Answers which then provide some development of the reason/s will begin to demonstrate some analysis skills and teacher/markers will then have to make a judgement in relation to the mark scheme about the quality of the analysis they are reading. This ought to mean the range of marks awarded is more limited.

A levels of response mark scheme requires teacher/markers to exercise some judgement about the quality of the answer and the extent to which the

student has demonstrated the assessment objectives being targeted. Some indicative content provides guidance as to the type of response that might be expected. Some possible arguments for and against London hosting the Games are provided along with a reminder that the focus of the question is to target the higher order skills of analysis and evaluation and that there is no 'right answer' to the question as far as the teacher/marker is concerned. It is the quality of the argument presented and how these higher order skills are demonstrated that helps determine the marks awarded.

Markers are given four levels each of which has some descriptive text showing what students may demonstrate in each level. The question is devised to target and encourage students to demonstrate the skills of analysis and evaluation (AO3). Each level has some reference to how students may demonstrate these skills. At level one, students may make judgements but these will be simplistic and lack any support. 'I think it is good for London to *host the Olympic Games'* being an example of such a simple judgement. Better quality answers will demonstrate analysis and evaluation skills at different levels; in level 2, a judgement may be given with some support but will tend to be one-sided whereas at level 3 the student is able to offer more points in support of their judgement and also recognise that there may be an opposing viewpoint. In Level 4 the student will be able to offer a more balanced argument and possibly use the 'it depends rule'; this is where students would recognise that the answer to the question may be dependent on different circumstances and perspectives. In reading through responses, markers have to make a judgement about the answer to decide, on the basis of the evidence, what level the answer sits in. Having decided on the level, the teacher/marker has to determine whether the quality of the analysis and evaluation is such that it warrants being at the top of the level, the bottom or nearer the middle.

5.2.2 The SOLO Mark Scheme

The preparation of markers for the SOLO mark scheme required different considerations from that of the Bloom scheme. The main reason, as has been

referred to above, is that many teachers in England and Wales may not be as familiar with this taxonomy. SOLO (Biggs and Collis 1982, Biggs and Tang 2007) was developed in relation to research being carried out in higher education. As a result, much of the language used in the descriptors is appropriate for that level of study rather than GCSE. This meant that I needed to develop the mark scheme in a format that markers would more easily understand and be able to interpret.

I initially drafted an introduction to SOLO in a similar way to that of the Bloom taxonomy (see Appendix XIII). The aim was to outline the key difference between SOLO and Bloom in that the former described, through five levels, how a learner's performance grows in complexity when mastering academic tasks. This increasing complexity relates to quantitative change (the amount of detail in the student's answer) and qualitative change where the detail is being integrated into a structured pattern. Quantitative stages of learning occur first and then learning changes qualitatively.

As SOLO is assessing learning outcomes it was important to outline what is meant by 'learning outcome' in this context. This was followed by a further brief outline of the task for the marker. In writing the guide for the SOLO mark scheme I was conscious of the need to provide guidance for teacher/markers in its use and offer an introduction to SOLO without making the mark scheme too detailed and 'academic', which might have alienated some participants. I also felt that it was important to maintain some consistency in approach to both mark schemes to avoid the possibility of influencing teacher/markers' outcomes.

The use of different mark schemes with different types of marking process (points-based mark schemes, levels of response mark schemes and a mark scheme based on quality of learning outcome) served to highlight some of the problems that can arise when assumptions are made about the efficacy of teachers/markers' conceptions of learning and recognition of threshold concepts and troublesome knowledge.

5.3 Data Analysis Using Descriptive Statistics Assuming Parametric Data

One of the early results of the research was that an assumption of the data generated from this marking process being continuous and capable of being subject to parametric tests was open to question. In the first stage I analysed the marks using standard statistical measures of central tendency and measures of dispersion. The use of such measures is widespread across national examinations and assumes consistent data. This implies that manipulation presents an accurate rather than erroneous interpretation of the data (Field 2009). However, as has been noted, continuous data has the characteristic of taking any value in a range which can be accurately measured according to set standards - one person's understanding and interpretation of each number in the series is the same. Analysing numerical data using the mean, range, standard deviation and so on, is appropriate when using continuous data. The marks using the Bloom mark scheme fitted the parameter of providing data with a numerical score capable of arithmetic manipulation but the SOLO mark scheme results being in levels, whilst providing a number, had limitations in terms of arithmetic manipulation.

This initial analysis showed variations in the assessment of the student answers by markers as shown in Table 5.1.
Descriptive Statistics							
	Ν	Range	Minimum	Maximum	Mean	Std.	Variance
Student						Deviation	
ST1	20	7.00	3.00	10.00	7.2000	1.57614	2.484
ST2	20	5.00	1.00	6.00	3.4500	1.23438	1.524
ST3	20	6.00	.00	6.00	2.4000	1.75919	3.095
ST4	20	7.00	.00	7.00	4.6000	1.81804	3.305
ST5	20	5.00	.00	5.00	2.1000	1.51831	2.305
ST6	20	7.00	.00	7.00	4.6000	1.75919	3.095
ST7	20	4.00	.00	4.00	.7500	1.25132	1.566
ST8	20	5.00	2.00	7.00	3.9500	1.43178	2.050
Valid N	20						
(listwise)							

Table 5.1 Descriptive Statistics for Bloom Mark Scheme, Question 1

Table 5.1 shows that there was some agreement between the markers that Student 1's answer most closely resembled the requirements of the mark scheme and the student was the best of this sample with a mean mark of 7.2 out of 10. However, there was a considerable variation between the markers with a range of 7 separating the lowest and highest marks awarded for this student.

The weakest performing student was Student 7 with a mean mark of less than 1 (0.75) and a range of 4 marks separating the lowest and highest scoring marks. Assessor 16 and Assessor 11 awarded the same mark to Student 7 as one teacher/marker awarded to Student 1 (3 marks). For two markers to look at the two answers as being of the same quality and deserving of the same marks when such a range exists raises questions as to the reliability of interpretation of the mark scheme or the way the student answered the question or a combination of the two. Most teacher/markers thought that

Page | 181

Student 1's answer demonstrated higher order skills in line with the mark scheme whereas most teacher/markers identified Student 7 as being weak and demonstrating few higher order skills. Some teacher/markers clearly had problems distinguishing the two. Of the remaining students, 4 of the 8 students had mean marks between 3.45 and 4.6 out of 10 and so were considered by teacher/markers as answers reflecting similar levels of performance. However, the range was similarly varied at between 5 and 7 marks out of a total of 10 available to be awarded.

Table 5.2 uses the same measures to compare the outcome of the marks awarded by the 20 markers using the SOLO mark scheme. This reveals some areas of agreement but equally some areas of disagreement.

	Ν	Range	Minimum	Maximum	Mean	Std.	Variance
Student						Deviation	
ST1	20	2.00	3.00	5.00	4.0000	64889	421
011	20	2100	0100	0100		101007	
ST2	20	2.00	1.00	3.00	2.1000	.44721	.200
ST3	20	2.00	1.00	3.00	1.7000	.73270	.537
ST4	20	3.00	1.00	4.00	2.4500	.99868	.997
ST5	20	3.00	1.00	4.00	1.8000	1.00525	1.011
ST6	20	4.00	1.00	5.00	3.2000	.89443	.800
ST7	20	2.00	.00	2.00	0.8500	.58714	.345
ST8	20	2.00	2.00	4.00	3.0000	.91766	.842
Valid N	20						
(listwise)							

Table 5.2 Descriptive Statistics for SOLO Mark Scheme, Question 1

Decorinting Statistics

As with the Bloom Taxonomy, Student 1 is considered the strongest student overall but there was some disagreement between teacher/markers over which level to put this student in, with a range of 2 from Level 3 to Level 5 Page | 182 being selected. Teacher/markers rated Student 7 as the weakest student again, as was the case when the Bloom Taxonomy, with a range of 2 (Level 0 – Level 2. There was, however, considerable disagreement over Student 6 with a range of 4; some teacher/markers rated this Student's answer in the top level (Level 5) whilst others rated the Student's answer in Level 1. Students' 4 and 5 answers also had wide ranges of 3 between Level 1 and Level 4.

Taking the mean scores, the ranking of the students' answers for both mark schemes can be given as shown in Table 5.3.

Rank	Student (Bloom)	Rank	Student (SOLO)
1st	1	1st	1
=2nd	4	2nd	6
=2nd	6	3rd	8
4th	8	4th	4
5th	2	5th	2
6th	3	6th	5
7th	5	7th	3
8th	7	8th	7

Table 5.3	Comparing	Ranking of Stud	dent Answers	to Question	1 using the
Bloom Tax	conomy and	the SOLO Taxo	nomy.		-

Whilst there appears to be agreement that Students 1 and 7 were the strongest and weakest using both mark schemes, there was disagreement over the ranking of Students 3, 4, 5, 6 and 8 although in both cases Students 4, 6 and 8 were in the top 4 whilst Students 3 and 5 vied over 6th and 7th.

In examinations, rank ordering is an important part of the overall assessment process. Standardisation should mean that markers would agree on the rank order of students in terms of evidence of learning regardless of the marks awarded but the variations in rank ordering shown by Table 5.3 suggests that a different mark scheme provides different interpretations of such a rank ordering and performance by this sample of teacher/markers.

Table 5.4 shows descriptive statistics for Question 2 using the Bloom mark scheme suggesting a similar degree of variation in assessing student performance.

Descriptive Statistics							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
ST1	20	6.00	6.00	12.00	9.4500	1.60509	2.576
ST2	20	10.00	6.00	16.00	9.0500	2.60516	6.787
ST3	20	4.00	6.00	10.00	8.0000	1.25656	1.579
ST4	20	7.00	6.00	13.00	10.6500	1.95408	3.818
ST5	20	9.00	3.00	12.00	7.8000	2.58742	6.695
Valid N	20						
(listwise)							

Table 5.4 Descriptive statistics for Dibuin Mark Scheme, Question 2

Student 4 was ranked as being the best with a mean mark of 10.65 out of 16 but there was a considerable range amongst the 20 markers with the lowest score given to this student being 6 out of 16 and the highest 13 out of 16. The range for other students was also varied with the lowest range being 4 for Student 3 and the highest at 10 for Student 2. There was clearly considerable disagreement amongst teacher/markers over the answer given to the question by Student 2 in relation to the mark scheme.

Table 5.5 shows the descriptive statistics for answers to question 2 using the SOLO mark scheme.

Descriptive Statistics							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
ST1	20	2.00	2.00	4.00	3.1500	.74516	.555
ST2	20	2.00	3.00	5.00	3.5500	.60481	.366
ST3	20	2.00	2.00	4.00	3.2000	.61559	.379
ST4	20	4.00	1.00	5.00	3.8500	.93330	.871
ST5	20	4.00	1.00	5.00	3.5500	.94451	.892
Valid N (listwise)	20						

Table 5.5 Descriptive Statistics for SOLO Mark Scheme, Question 2

Table 5.5 highlights similar disagreements over the quality of the answers using the SOLO mark scheme. The mean marks were much closer but this is due to the smaller range of available levels. The standard deviation and variance for each student are also much smaller than those for the Bloom mark scheme as a result but the range suggests some disagreement. The range varied from 2 to 4 across the 5 students. Student 4 was ranked the best in agreement with the interpretation of the Bloom mark scheme. However, the range for this student's answers was 4 with some markers putting Student 4 into Level 1 and others into Level 5.

These differences in interpretation of the mark scheme for the same answers is highlighted through the use of these basic statistical tests. If there are differences in interpretation of answers then the appropriateness of using descriptive statistics to analyse this data is called into question. I looked for some further insight through some of the comments made by teacher/markers on the papers they returned.

Comments from 8 teacher/markers in relation to Student 2 are provided below. The teacher/markers were using the Bloom mark scheme and some chose to make comments to help justify their marks. On reading some of the Page | 185 accompanying comments on the student answers written by some teacher/markers, it became further apparent that the assumption of parametric data in this sample could not be taken for granted.

Of these 8 markers, half agreed on a mark of 8 for Student 2 and 6 agreed that the answer included two sides of the argument. However, there was clear disagreement on whether there was any judgement included in the answer. Assessor 20 notes that there is no judgement formed as did Assessor 18 and whilst they agreed on this point Assessor 18 awarded the answer 8 marks but Assessor 15 only 6 marks. Assessor 15 thought that the analysis was not developed but Assessor 20 saw "good analysis" and Assessor 1 saw some "well developed analysis". Assessor 4 seemed to see some skills that some other teacher/markers did not see, noting that the answer showed "good application and use of business theory" and that "both sides of the argument are analysed and evaluated". Assessors 1 and 2 also saw some higher level skills placing the answer in Level 3.

From these comments it seems clear that different markers are seeing exactly the same answer in a different light and interpreting evidence of learning and skills in the answer in different ways. Even where there is an agreement on the level of response in the answer there is disagreement about where the answer lies in that level and on the reasons for awarding a particular mark.

Comments made by markers of Student 5 using the SOLO mark scheme are given in Table 5.6. Using the SOLO mark scheme with Student 5's answer, the 'marks' ranged from Level 1 to Level 4. The quality of student answers in these different levels is considerable and again it seems that markers are looking at the same responses in a very different way.

Assessor	Comment
Assessor 1	"The answer could be Band 4 as it includes a judgment from the outset. The advantages and disadvantages are weighted and related to new terms, and concepts such as GDP are mentioned. More than two linked factors are shown. My Concern is the weak conclusion. I have classed it as Band 3 as the conclusion replicates the initial view and does not make use of some of the facts to make a more

	effective judgment from the evidence given".
Assessor 2	"Band 4 - Bullets all met, although a great deal more evidence could be drawn upon to link to concepts and theories. Underdeveloped points. Not as sophisticated an answer as candidate 4".
Assessor 4	"Band 5/4 - Relevant knowledge and theory applied to the answer seamlessly with learner not merely demonstrating volume of knowledge. Logical and selective evidence extracted from the support material to support statements, with a coherent structure and 1 sided personal opinionated conclusion as learner has not provided an objective judgment".
Assessor 7	"Band 4 - Good development and linking of concepts but not beyond usual teaching".
Assessor 8	"Band 3 - Brief but enough for this band at least".
Assessor 10	"Band 4. A concise answer with some points demonstrating their own thinking. However a lack of a justified judgement / balanced conclusion meant it didn't reach the next level".
Assessor 18	"Band 4 [BOD (benefit of doubt)] - makes sensible use of evidence and tries to assess, weak conclusion".
Assessor 19	"Quality of the argument is weak and is based on a limited amount of data - I am not sure the conclusion follows from the evidence presented? Band 1".
Assessor 20	"Band 3 - No significant analysis. Simple judgements only".

Assessor 19 notes that the "quality of the argument is weak" and is not convinced of the logic of the conclusion, placing the answer into Level 1. Assessor 2 notes that the points made in the answer were "undeveloped" but still puts the answer in Level 4 and suggests "all the bullets have been met" even though bullet two in the scheme requires evidence of an 'understanding of several issues/elements/components which are conceptually integrated'. Assessor 7 by contrast sees "Good development and linking of concepts" and Assessor 4 sees "Logical and selective evidence extracted from the support material to support statements, with a coherent structure...".

Given these very different interpretations, questions have to be raised about what each teacher/marker in this sample 'understands' about the mark scheme, the levels within and the marks they are awarding. The implication is that teacher/markers in this sample have a different understanding of what evaluation means and what constitutes a judgement. If there is a different interpretation of the mark schemes and the marks/levels being awarded then the data is not continuous but categorical.

The use of descriptive statistics for analysing this data may not, therefore, be appropriate. Each marker is not interpreting the mark scheme and the numerical values in the same way. The interpretation of the student answers suggested by the ranges in the data suggests that the answers mean different things to different markers even though they are using the same mark scheme and are reading the same answer. Rather than being continuous, the data appeared to more closely resemble categorical data, which necessitates different statistical approaches.

5.4 Z-Scores

If the data is categorical rather than continuous, non-parametric statistical tests can be used to see the extent to which teacher/markers agree and the probability that the differences in interpretation are simply random or have some more specific explanation.

One of the problems the use of two different mark schemes generated is the difficulty of drawing a comparison between the two given the fact that they both used different 'scores' (the Bloom mark scheme using marks out of 10 and 16 and the SOLO using Levels). Trying to compare the reliability of teacher/markers interpretation when using these different base criteria presented a challenge. A way of overcoming this to standardise the two mark schemes to enable comparison. One method which can be adopted is the use of *z*-scores, A *z*-score describes the location of the marks in the distribution (Gravetter and Wallnau, 2011) and inform a standardised distribution that can be directly compared to other distributions that have also been transformed into *z*-scores (e.g. comparing the marks awarded using Bloom to those awarded using SOLO).

The formula for finding z-scores is:

$$Z = \frac{x - \mu}{\sigma}$$

where x is the score, μ the mean and σ the standard deviation. Given the Page | 188

data for Bloom (20 markers, 8 students and 160 marks) I calculated the total marks awarded for the 8 students by the 20 markers. Using the Bloom scheme this figure is 583. There are 160 marks so the average mark (μ) is 3.64 and the standard deviation (σ) is 2.32. To calculate Student 1's *z*-score, therefore, I followed the process below:

Student 1's z-score is their mean (7.2) minus μ (3.64) = 3.56 / σ (2.32) = 1.53

Student 2's z-score is similarly found thus: 3.45 - 3.64/2.32 = -0.08

This same approach can be applied using the SOLO mark scheme:

Total 'marks' awarded = 382, μ = 2.39, σ = 1.22

Student 1's z-score will be 4 - 2.39/1.22 = 1.32

Student 2's z-score will be 2.1 - 2.39/1.22 = -0.23

By calculating the *z*-scores, this tells me that using Bloom, markers locate Student 1 above the mean by about 1.5 standard deviations whereas using SOLO they also locate Student 1 above the mean but only by one and a third standard deviations. Similarly, using Bloom, markers locate Student 2 below the mean by less than a tenth standard deviations but using SOLO, whilst they also locate Student 2 below the mean, they locate the student almost a quarter standard deviations below. What this shows is that there is some comparability in terms of location above and below the mean but greater variability in where it is located above and below the mean.

This process was repeated for all the students across the two mark schemes for each question. Transforming the raw scores for both mark schemes into *z*scores and then transforming these into 'marks' gave a standardised distribution allowing comparison for each student. For example, if Student 1 was awarded an average of 7.2 using Bloom with a μ of 3.64 and σ = 2.32, and a mean of 4 using SOLO with μ = 2.39 and σ = 1.22, I can then translate these marks into a standardised mark which allows some comparison through a standardised μ and σ .

Student	Mean	Mean	z-score	z-score
	(Bloom)	(SOLO)	(Bloom)	(SOLO)
Student 1	7.20	4.00	1.49	1.32
Student 2	3.45	2.05	-0.08	-0.28
Student 3	2.40	1.70	-0.51	-0.57
Student 4	4.60	2.42	0.41	0.03
Student 5	2.10	1.79	-0.64	-0.49
Student 6	4.60	3.20	0.41	0.66
Student 7	0.75	0.85	-1.21	-1.26
Student 8	3.95	3.00	0.13	0.50

Table 5.7Z- score Calculation for Question 1 Using Bloom and SOLO MarkScheme

Table 5.7 shows that the z-scores for each student reveal differences in the way in which student answers have been interpreted by markers across the two mark schemes. Student 1, for example, is above the mean on both mark schemes. Using the Bloom mark scheme, the z-score shows a variation of almost one and a half standard deviations and almost one and a third standard deviations using SOLO. Four students are above and below the means using both mark schemes which indicates some degree of comparability but with some variation.

To take the analysis one step further I used the z-score data to create a standardised distribution with new values for μ and σ which will allow more obvious comparison without changing the location of each student in the distribution. I assumed a standardised $\mu = 50$ and $\sigma = 10$ (these values could be anything provided they are applied consistently across the distributions). Recall that Student 1 has a *z*-score of 1.49 using the Bloom mark scheme and 1.32 using SOLO.

Referring back to the formula:

$$Z = \frac{x - \mu}{\sigma}$$

I can substitute the relevant scores into the formula to get standardised scores.

For student 1:

$$1.49 = (x - 50)/10$$
$$14.90 = x - 50$$

Adding 50 to both sides gives:

$$14.90 + 50 = x$$

 $x = 64.90$

For Student 1 using the SOLO mark scheme:

$$1.32 = (x - 50)/10$$
$$13.20 = x - 50$$
$$13.20 + 50 = x$$
$$x = 63.20$$

This suggests that the marks awarded to Student 1 using the Bloom mark scheme were slightly higher at 64.90 compared to 63.20 with a difference of 1.7. This difference can be seen as being negligible.

Repeating this process for the other students for question 1 gives the outcome shown in Table 5.8:

Table 5.8	Standardised	z-scores for	Question 1	١.
-----------	--------------	--------------	-------------------	----

Question 1	STD scores (Bloom)	STD scores (SOLO)	% Difference (Bloom over SOLO)
Student 1	64.94	63.20	+2.75
Student 2	49.25	47.23	+4.27
Student 3	44.85	44.34	+1.10

Page | 191

Student 4	54.06	50.25	+7.58
Student 5	43.60	45.08	-3.39
Student 6	54.06	56.64	-1.07
Student 7	37.95	37.38	+1.52
Student 8	51.34	55.00	-7.12

Student 1 is the highest scoring student on both schemes with a difference in marks of +2.75% using the Bloom scheme. There is also agreement that Student 7 is the weakest with the marks being very similar with just a 1.52% difference. The difference in marks for Student 8 is over 7% as is the mark for student 4. Student 8 is marked more highly on the SOLO mark scheme (-7.12%) and Student 4 more highly on the Bloom mark scheme (+7.58%). In five of the observations the marks awarded using the Bloom scheme are higher than those of the SOLO scheme (indicated by a plus sign).

Table 5.9 shows that differences in the marking between the two mark schemes also changes the rank ordering. Students 4 and 6 are ranked joint second using the Bloom mark scheme but the SOLO scheme has Student 6 ranked second and Student 8 third compared to equal second and fourth respectively using the Bloom mark scheme. Students 3 and 5 swap places in the rankings at sixth and seventh.

Rank order (Bloom)	Student	STD scores (Bloom)	Rank order (SOLO)	Student	STD scores (SOLO)
1	Student 1	64.94	1	Student 1	63.20
=2	Student 4	54.06	2	Student 6	56.64
=2	Student 6	54.06	3	Student 8	55.00
4	Student 8	51.34	4	Student 4	50.25

Table 5.9 Rank Order: Question 1.

5	Student 2	49.25	5	Student 2	47.23
6	Student 3	44.85	6	Student 5	45.08
7	Student 5	43.60	7	Student 3	44.34
8	Student 7	37.95	8	Student 7	37.38

Repeating the process for Question 2 gives the outcome shown in Table 5.10:

	STD scores (Bloom)	STD scores (SOLO)	% Difference (Bloom over SOLO)
Student 1	52.56	46.17	+13.80
Student 2	48.45	51.11	-5.49
Student 3	45.94	46.79	-1.85
Student 4	58.04	54.81	+5.89
Student 5	45.02	51.11	-13.53

 Table 5.10 Standardised z-scores for Question 2.

Table 5.10 shows some notable differences in the marks awarded to students using the SOLO mark scheme. In this case Student 1 received a mark that was 13.8% higher using the Bloom mark scheme whereas Student 5 received a mark using SOLO 13.53% lower than the Bloom mark scheme. The standardised marks across both schemes are relatively low with nothing higher than 58.04 being awarded.

The rank ordering for Question 2 is also different as shown in Table 5.11.

Rank order	Student	STD	Rank order	Student	STD scores
(Bloom)		scores	(SOLO		(SOLO)
		(Bloom)			
1	Student 4	58.04	1	Student 4	54.81
2	Student 1	52.56	=2	Student 2	51.11

Table 5.11 Rank Order: Question 2.

Page | 193

3	Student 2	48.45	=2	Student 5	51.11
4	Student 3	45.94	4	Student 3	46.79
5	Student 5	45.02	5	Student 1	46.17

There was agreement that Student 4 was the strongest using both schemes but Student 1 was considered the weakest on the SOLO scheme but ranked second using the Bloom mark scheme. Student 5 was considered the weakest using the Bloom scheme but ranked equal second using SOLO.

Z-scores are predicated on the distribution of the sample being normal. It was important to assess, therefore, the extent to which the data I am working with represents a normal distribution to allow valid conclusions to be drawn from the above test and to see if this approach represents a better fit for comparing the two mark schemes.

5.4.1 Tests for Normality

If (*p*>.05) then the distribution of the sample is not significantly different to a normal distribution (it is deemed 'probably' normal) but if (*p*<.05) the distribution is significantly different to a normal distribution - it is non-normal. To run tests of normality on the data, I used SPSS. To run the tests for normality I selected '**Analyze> Descriptive Statistics >Explore**...'. I then transferred the variable that needs to be tested for normality into the 'Dependent List' box. I then selected the 'Plots' button. I selected 'Normality plots with tests'. This produces the Kolmogorov-Smirnov and Shapiro-Wilk tests as shown in Table 5.12. These tests compare scores in a sample to a set of scores with a normal distribution having the same mean and standard deviation (Field, 2009). If the **Sig.** value of the Shapiro-Wilk Test is greater than 0.05 then the data is normal, i.e. the distribution of the sample is not significantly different from a normal distribution (Field, 2009). If it is below 0.05 then the data significantly deviates from a normal distribution.

Tests of Normality							
	Kolmogorov	/-Smirnov ^a		Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
ST1	.194	20	.047	.928	20	.140	
ST2	.222	20	.011	.935	20	.189	
ST3	.190	20	.057	.906	20	.052	
ST4	.137	20	.200 [*]	.934	20	.186	
ST5	.266	20	.001	.835	20	.003	
ST6	.237	20	.005	.881	20	.018	
ST7	.376	20	.000	.664	20	.000	
ST8	.186	20	.068	.927	20	.132	
a. Lilliefors Significance Correction							
*. This	is a lower b	oound of t	he true sig	nificance.			

Table 5.12 Bloom Q1 Normality Test

For Question 1 using Bloom the K-S test reveals that in 3 of 8 students, D (20) = p < .05 indicating that the distribution is non-normal (the scores presented in red).

Table 5.13 SOLO Q1 Normality Te

Tests of Normality							
	Kolmogorov	/-Smirnov ^a	I	Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
ST1	.300	20	.000	.793	20	.001	
ST2	.438	20	.000	.611	20	.000	
ST3	.280	20	.000	.784	20	.000	

ST4	.209	20	.022	.887	20	.024
ST5	.287	20	.000	.770	20	.000
ST6	.262	20	.001	.881	20	.019
ST7	.351	20	.000	.754	20	.000
ST8	.262	20	.001	.752	20	.000
a. Lilliefors Significance Correction						

For Question 1 using the SOLO mark scheme, the K-S test shows all 8 students D (20) = p < .05, indicating that the distribution is non-normal.

Table 5.14 Bloom Q2 Normality Test

Tests of Normality							
	Kolmogorov	/-Smirnov ^a	l	Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
ST1	.190	20	.058	.944	20	.281	
ST2	.357	20	.000	.814	20	.001	
ST3	.200	20	.035	.910	20	.063	
ST4	.171	20	.127	.911	20	.068	
ST5	.179	20	.094	.932	20	.170	
a. Lilli	efors Signifi	cance Cor	rection				

For Question 2 using the Bloom mark scheme, the K-S test shows 1 of 5 students D (20) = p< .05 (data in red) indicating that the distribution is non-normal.

Table 5.15 SOLO Q2 Normality Test

Tests of Normality						
	Kolmogorov	v-Smirnov ^a		Shapiro-Wi	lk	
	Statistic	df	Sig.	Statistic	df	Sig.
ST1	.230	20	.007	.809	20	.001
ST2	.318	20	.000	.737	20	.000

Page | 196

ST3	.327	20	.000	.771	20	.000
ST4	.414	20	.000	.686	20	.000
ST5	.230	20	.007	.847	20	.005
a. Lillio	efors Signifi	cance Cor	rection			

For Question 2 using the SOLO mark scheme, the K-S test shows all 5 students D (20) = p < .05 indicating that the distribution is non-normal.

The test results for the Bloom mark scheme overall show 4 out of 13 students with distributions that are likely to be non-normal whereas for the SOLO mark scheme all 13 students have Sig. outcomes that suggest the distribution is non-normal. Given the test results, it is concluded that it is not safe enough to view the distribution across both mark schemes, and with both questions, as being normal. This suggests that the use of z-scores and standardised distributions would not represent a good fit despite highlighting some variability in the approach of markers to student answers using the different mark schemes.

5.5 Coefficient of Variation

Since it was not possible to rely on z-scores to indicate the difference between the two mark schemes, I decided to look at the coefficient of variation between the two mark schemes. Using the coefficient of variation will indicate the variation of marks using the Bloom mark scheme in comparison to the variation in marks using the SOLO mark scheme. Specifically, the coefficient of variation indicates whether the variation in marks is greater than or less than those using the SOLO mark scheme.

The coefficient of variation of a set of observations is the standard deviation of the observations divided by their mean (Gravetter and Wallnau, 2011). It indicates that the sample standard deviation is x % of the value of the sample mean. The size of the standard deviation is dependent to some extent on the sample size, for example, in a sample of 1 million, a standard deviation of 10

might be seen as small but in a sample size of 50 would be considered large. In this instance the sample size is small. If the coefficient of variation is <0.25 (25%) this is associated with low variability; a coefficient of variation between 0.25 and 0.75 (25% and 75%) is associated with moderate variability and a coefficient of greater than 0.75 (75%) with a high degree of variability.

Table 5.16 shows the coefficient of variation for each of the 20 markers for both the Bloom mark scheme and the SOLO mark scheme.

Marker	Bloom Q1 CoV (%)	SOLO Q1 CoV (%)
Assessor 1	60.5	47.1
Assessor 2	74.9	47.8
Assessor 3	52.9	43.4
Assessor 4	70.9	47.8
Assessor 5	38.8	37.0
Assessor 6	48.3	25.2
Assessor 7	77.5	34.9
Assessor 8	81.0	65.5
Assessor 9	62.7	53.0
Assessor 10	50.0	57.0
Assessor 11	57.1	63.4
Assessor 12	93.2	54.8
Assessor 13	77.3	72.3
Assessor 14	67.3	50.7
Assessor 15	57.1	50.0
Assessor 16	26.8	25.2
Assessor 17	61.4	58.7
Assessor 18	95.9	59.8

Table 5.16	Coefficient	of Variation	Ouestion	1.
	oberneient	or variation	Question	•••

Page | 198

Assessor 19	73.1	77.3
Assessor 20	116.8	73.1

None of the scores are below 25% suggesting that at the very least there is moderate variability in the marks between the two mark schemes. In the Bloom mark scheme, 14 of the 20 coefficients are in the range 25% to 75% with 11 of these 14 having coefficients above 50%. The remaining 6 coefficients are above 75%. It can be concluded that overall there is a high degree of variation using the Bloom mark scheme.

For the SOLO mark scheme 19 of the 20 coefficients fall in the range 25% to 75% with 9 of these being at or below 50%. The pattern of the coefficients between the two mark schemes is highlighted in Figure 5.1.



Figure 5.1 Coefficient of Variation for Question 1

The closer the red square to the blue diamond the more similar is the response of the teacher/marker regardless of mark scheme. It can be seen that assessors 5, 10,11, 13, 15, 16 and 17 have marks related to both mark schemes which are very similar but the remaining 13 teacher/markers show some evidence of variance in the way they interpret the same student answer according to the mark scheme they are using. Assessors 2, 4, 6, 7, 12, 18 and Page | 199

20 appear to have a noticeable difference between the interpretations of the student answer using the two mark schemes.

Table 5.17 shows the coefficient of variation for Question 2.

Marker	Bloom Q2 CoV	SOLO Q2 CoV
Assessor 1	19.6	15.2
Assessor 2	15.2	17.7
Assessor 3	21.1	16.1
Assessor 4	38.2	25.0
Assessor 5	35.3	10.6
Assessor 6	51.1	11.8
Assessor 7	17.8	15.2
Assessor 8	21.9	14.0
Assessor 9	22.1	14.0
Assessor 10	12.1	29.9
Assessor 11	20.0	0.00
Assessor 12	18.7	31.7
Assessor 13	28.3	22.0
Assessor 14	16.1	16.1
Assessor 15	24.1	38.0
Assessor 16	9.5	15.2
Assessor 17	17.4	14.0
Assessor 18	28.6	33.3
Assessor 19	44.8	46.4
Assessor 20	24.8	16.1

 Table 5.17 Coefficient of Variation Question 2.

Table 5.17 suggests there is a lower variability in both the Bloom and SOLO marks schemes for this question. For both mark schemes, 14 out of 20 markers had a coefficient of variation <0.25 indicating low variability and the remaining 6 markers were all between 0.25 and 0.75 suggesting only medium variability. Whilst there is a lower overall degree of variability between markers with Question 2 using the two mark schemes, there are some noticeable differences in individuals when looking at the representation of the table as in Figure 5.2.



Figure 5.2 Coefficient of Variation for Question 2

Assessors 5, 6, 10, 11 and 15 have coefficients that are very different when looking at how the student answer was interpreted using the different mark schemes.

I applied the coefficient of variation at student level for each question as shown in Table 5.18 and Figure 5.3.

Student	Bloom Q1	SOLO Q1
4	21.0	1()
1	21.9	16.2
2	35.8	21.3
3	73.3	43.1
4	39.5	40.8
5	72.3	55.8
6	38.2	28.0
7	166.8	69.1
8	36.2	30.6

Table 5.18 Coefficient of Variation by Student, Question 1

Table 5.18 shows that looking at the student level using the Bloom mark scheme only Student 1 has a coefficient below 0.25. Six students have a coefficient of variation between 0.25 and 0.75, although of these, four are lower than 0.5 suggesting the degree of variation is relatively low. That leaves 1 student with a coefficient of variation greater than 0.75 and it can be seen that the coefficient of variation for student 7 is extremely high at 166.8%.

For the SOLO mark scheme, 2 students have a coefficient of variation lower than 0.25 and 6 have a coefficient of variation between 0.25 and 0.75. Overall the conclusion is that some degree of variability in the marking exists for question 1.



Figure 5.3 Coefficient of Variation by Student, Question 1

Figure 5.3 represents this information in chart form and shows that there is some agreement on the marks for students 1, 4, 6 and 8 between the two mark schemes but that there is some variability between the remianing students with the degree of variability between the mark schemes for Student 7 being particularly marked. In this case the Bloom mark scheme has a higher set of coefficients in 7 of the 8 students.

Student	Bloom Q2	SOLO Q2	
1	17.0	23.7	
2	24.7	29.5	
3	15.7	19.2	
4	18.3	24.2	
5	33.2	35.6	

Table 5.19	Coefficient	of Variation	by Student.	Ouestion 2
	cochicicit		by student,	Question 2

Table 5.19 shows the outcome for question 2 by student. In this case four of the five students have a coefficient of variation for the Bloom mark scheme

below 0.25 and the one that is between 0.25 and 0.75 is towards the lower end of the scale indicating low levels of variability. Three of the five students in relation to the SOLO mark scheme have low levels of variability and the remaining two have moderate variability with a coefficient between 0.25 and 0.75.



Figure 5.4 Coefficient of Variation by Student, Question 2

Figure 5.4 shows a greater degree of comparability between the two mark schemes although in each case the variability is higher when looking at the SOLO mark scheme.

Overall the use of the coefficient of variation shows that there is some evidence of moderate to large variability in the markers using the different mark schemes for Question 1 but less so for Question 2. This would imply that there is some factor which is leading to a difference in the way the mark schemes are being used and that the null hypothesis can be rejected.

5.6 Non-parametric Tests - Chi-square

Non-parametric tests can be used to overcome the problem of using descriptive statistics to analyse categorical data. From the analysis so far, there is some evidence that the mark scheme used does have some effect on the interpretation of student answers by markers. Looking for a test that statistically proves the degree of significance is compromised by the different

nature of the two mark schemes. Assessing the degree of independence between two sets of categorical data can be identified with the use of chisquare.

The test of independence is performed to determine whether the two variables in a contingency table (a way of displaying data so that the relationship between categorical variables can be established) are independent of each other or not when a paired observation is conducted (Gravetter and Wallnau, 2011). The null hypothesis remains that there is no difference to the way a marker responds to a student answer in relation to the mark scheme used. Chi-square tests are based on frequencies and the difference between the frequency of the observed event and the forecast or expected event. It is possible to take account of the difference in the mark schemes to set up a contingency table to run chi-square. This can be done by setting up a grid as in Table 5.20 which allows the scores to be converted for both mark schemes and both questions to a common scale with 6 levels in each with each level linked to a letter, A – F. These letters are given a number ranging from 10 – 60 so that all of the marks for both questions and for both mark schemes could now be expressed as a common value.

SOLO Mark Scheme Mark Conversion Table
Level 0 = F - 10
Level 1 = E - 20
Level 2 = D - 30
Level 3 = C - 40
Level 4 = B - 50
Level 5 = A - 60

For Bloom the marks are arranged as shown in Table 5.21.

Table 5.21 Bloom Mark Scheme Mark Conversion

Bloom Mark Scheme Mark Conversion Table
0 marks = F - 1
1 - 2 marks = E - 20
3 - 4 marks = D - 30
5 - 6 marks = C - 40
7 - 8 marks = B - 50
9 - 10 marks = A - 60

Having interpreted the data in this way, a contingency table can be drawn up such as the one shown in Table 5.22.

Scores\Taxonomy	Bloom (Assessor 1)	Solo (Assessor 1)
60	Frequency or counts	Frequency or counts
50	Frequency or counts	Frequency or counts
40	Frequency or counts	Frequency or counts
30	Frequency or counts	Frequency or counts
20	Frequency or counts	Frequency or counts
10	Frequency or counts	Frequency or counts

Table 5.22	Example of a	Contingency Table
------------	--------------	-------------------

This enables the observation of the frequency with which a particular mark was awarded by each marker across the two mark schemes. Table 5.23 demonstrates one example using Assessor 1:

Scores/Taxonomy	Assessor 1 (Bloom)		Assessor 1 (SOLO)	
60		0		1

Page | 206

50	0	3
40	1	0
30	1	3
20	1	1
10	6	0
0	0	0

Table 5.23 shows that Assessor 1 awarded the 'mark' of 10, 6 times using the Bloom mark scheme but only once using the SOLO mark scheme. Assessor 1 awarded the highest marks of 50 and 60 a total of 4 times using the SOLO mark scheme but none using the Bloom scheme. This process could be repeated for all the teacher/markers. This would suggest that there is some effect from the use of a different mark scheme but this intuitive inspection is insufficient to judge the degree of independence.

In order to use chi-square certain parameters need to be in place. Unlike some types of categorical data where there will be some response with which to measure the frequency, such as 'yes' or 'no' as responses to a question, not awarding a mark in the range available means that the contingency table will have a number of zeros indicating that the marker has not awarded that mark. There are 7 columns in the contingency table ranging from 0 – 10. The degrees of freedom will be affected because there are no values for the frequencies of certain grades, i.e. where the marker did not award that mark. In table 5.23 above, Assessor 1 did not award any marks in the 0 or 10 categories using either mark scheme. As a result the degrees of freedom will be calculated on the basis of five columns where a frequency is recorded rather than 7. The degrees of freedom will be: df=(5-1)(2-1)=4.

Field (2009), notes that to use the chi-square test the expected frequencies in each cell must be greater than 5. When the expected frequencies are too low, it probably means that the sample size is too small and that the sampling distribution of the test statistic is too deviant from a chi-square distribution Page | 207 to be of any use. In this instance the expected frequencies for each cell are less than 5.

Running a chi-square test using the data in Table 5.23 provides an output as shown in Table 5.24.

Assessor 1 Bloom * Assessor 1 SOLO Cross tabulation						
		Assessor 1 SOLO			Total	
			.00	1.00	3.00	
Assessor 1 Bloom	.00	Count	2	1	1	4
		Expected Count	1.7	1.1	1.1	4.0
		% of Total	28.6%	14.3%	14.3%	57.1%
	1.00	Count	1	0	1	2
		Expected Count	.9	.6	.6	2.0
		% of Total	14.3%	.0%	14.3%	28.6%
	6.00	Count	0	1	0	1
		Expected Count	.4	.3	.3	1.0
		% of Total	.0%	14.3%	.0%	14.3%
Total		Count	3	2	2	7
		Expected Count	3.0	2.0	2.0	7.0
		% of Total	42.9%	28.6%	28.6%	100.0%

Table 5.24 Assessor 1 Chi-square Output

Chi-Square Tests				
	Value	df	Asymp. Sig. (2-sided)	
Pearson Chi-Square	3.500 ^a	4	.478	
Likelihood Ratio	4.016	4	.404	
Linear-by-Linear Association	.000	1	.984	
N of Valid Cases	7			
a. 9 cells (100.0%) have expected count less than 5. The minimum				
expected count is .29.				

Table 5.24 confirms that the expected count in each table is less than 5 and as a result it is concluded that using chi-square as a test of independence would be flawed given the size of the data.

There are circumstances where sparsely populated cells in a contingency table is acceptable. This can be the case where it is impossible to estimate in advance the cell values (Kirkman, 1996). Prior to sending out the student answers and mark schemes it was not possible to predict what marks would be awarded to students and what the range might be – indeed, this was exactly what the investigation was designed to try and find out. Yarnold (1970) states that the minimum expectation can be less than one when the proportion of small expectation cells is also small, (Kirkman, 1996).

Despite these possibilities, the use of chi-square in this research is not possible because there is a further problem with using chi-square in this instance. Field (2009, p691) notes: "For the chi-square test to be meaningful, it is imperative that each person, item or entity contributes to only one cell of the contingency table. Therefore, you cannot use a chi-square test in a repeated measures design". As noted, this phase uses a form of repeated measures design where individuals are providing an interpretation of a student's answer in relation to one mark scheme and then at some later point, using the same student answer but assessing it using a different mark scheme. As a result, the use of chi-square to test independence was not Page | 209 deemed viable.

5.7 Summary and Conclusion

This phase of the research has shown that drawing comparisons between two mark schemes that focus on different aspects of evidence of learning presents problems in finding an appropriate statistical test for significance. Using parametric data was identified as being appropriate based on the fact that such statistical methods are used throughout England and Wales in national examination assessments. The assumption that different markers will understand the mark schemes and be in a position to recognise evidence of learning could be called into question based on the evidence of this phase of the research. The extent to which this is generalizable across national examination assessments is difficult to judge. On the one hand, all the teacher/markers in this phase of research (accepting that it is a small sample) are involved with teaching the GCSE economics and business gualification and in some cases are examiners for different awarding bodies. It could be argued that these teacher/markers are representative of teachers and examiners involved with GCSE economics and business. On the other hand, those teachers that are involved with awarding bodies on an official basis are subject to a far greater degree of standardisation than was possible in this phase of the research. The sample used in this phase may not, accepting this limitation, be representative. What this phase of the research does show is that the teacher/markers in the sample exhibited multiple realities in the way in which they interpreted the student answers using both mark schemes.

In this phase of the study, analysis using parametric methods appears to be flawed because the assumptions of a common understanding of what constitutes evidence of learning and how mark schemes are interpreted is not present. This means that non-parametric methods have to be used. With the mark schemes used in this phase of the research being different in their approach, finding a test that confirms that there is some statistical significance in the way in which markers interpret mark schemes and assess evidence of learning is problematic. Comparing the interpretation using the same mark scheme would have been an easier task but would not have raised the issue of whether the adoption of a different approach to assessment can be subject to comparison in order to draw a conclusion as to which is more valid and reliable as an assessment device.

Given that there is some evidence that teacher/markers in this study do vary in the way they interpret evidence of learning given by student answers, using the same mark scheme and between different mark schemes, it is possible that their understanding of what learning means and how such learning can be assessed varies also. The next phase of the research will look at teachers' conceptions of learning. This phase of the research has highlighted that there are different interpretations of evidence of learning in student answers, the sort of task that teachers carry out every day.

6. Phase III: Results

6.1 Introduction

The research for this thesis is based on three separate but related phases exploring the extent to which students in economics and business at this level are being taught to think in the subject. To think in the subject requires teachers to provide an environment whereby deep learning is encouraged so that learners begin to change behaviour and have a transformed view of the world.

Phase I of the research involved text analysis of student answers to examination questions. The results of the text analysis suggested that despite having relatively high marks, there was evidence of troublesome knowledge and students using schemas in extended writing answers. This allows students to meet the assessment objectives and secure marks to get high grades but not necessarily demonstrating deep learning.

This led to the second phase of the research which investigated whether teachers have a collective sense of what learning in economics and business looks like in relation to evidence provided by student answers to exam-type questions. Phase II generated some evidence to suggest that markers saw the same student answers differently and that teacher/markers in the sample had different interpretations of the evidence of learning, not only within the mark schemes but also between the two different mark schemes, although it was not fully clear that there was any consistent difference in the marking.

Phase III seeks to obtain teachers' conceptions of learning in the context of GCSE economics and business and the extent to which teaching is based on encouraging deep learning compared to what actually happens in the classroom on a regular basis. Evidence from the first two phases might be reinforced by evidence in the third phase that teachers revert to teaching schema rather than deep learning and thinking in the subject, possibly because of perceived external pressures.

Teachers are central figures in the learning process. There has been

considerable research into students' conceptions of learning (Marton, Dall'Alba and Beaty, 1993; Meyer, 1995; Burnett, Boulton-Lewis and Campbell, 1996, Prosser, Trigwell, Hazel and Waterhouse, 2000; Dart et al, 2000), but more limited research on teachers' conceptions of learning (Donche, et al 2007, Sanneke et al, 2004) being two examples), especially at the school level. This thesis is looking at the problematic nature of examining GCSE economics and business. One element of this involves the conceptual foundations of the disciplines and the presence of a particular language and abstract concepts making the teaching and learning of the subject challenging. For students at this level to begin to have some transformed understanding and to take the first steps to thinking in the subject, it is likely that they will experience troublesome knowledge and teachers have to help students navigate through this troublesome knowledge. Teachers might have a conviction to want to encourage deep learning and a transformed understanding of the subject in their students but instead adopt an approach to learning which not only provides schemas but which focus more on surface or strategic learning in an attempt to help students get the grades that they need and by which teachers are increasingly judged in England and Wales.

6.2 Results

6.2.1 Question 1

Question 1 (see Appendix VIII) is designed to see the extent to which respondents in the sample agree about what is a definition of learning. Answers to the question can also reveal the extent to which respondents have a conception of learning which is reflective of a surface approach or a deep learning approach and can be used as a comparison to answers in later questions where consistency in the definition of learning by respondents can be judged.

There were 17 instances where respondents' answers reflected conception 1 (C1). 'Finding out knowledge', 'assimilating knowledge', 'knowing something', 'developing new or existing knowledge', 'gaining knowledge', 'acquiring knowledge', 'taking in information' and 'being able to recall' Page | 213 were typical phrases used across the 22 responses. One curious response from R18 simply said '*transforming information into knowledge*' which raises a question about what the difference is between information and knowledge and how one can be transformed into the other.

Respondents who offered more detail in their answers tended to be the ones where an element of conceptions 4 - 6 were implied. References were made to 'extending beyond their own experiences', 'gains an understanding of the world around them', 'demonstrating learning in different contexts', 'thinking about topics in a new way', 'a permanent acquisition of something new - not just a temporary change' and 'finding meaning'. Most of these responses can be categorised in conceptions 4 and 5 (C4 and C5). Only two respondents, however, (R11 and R19) explicitly stated a definition of learning which could be categorised as conception 6 (C6) with a reference to learning being a state where 'how we think or act is changed' and '...results in an individual being able to think about a topic, problem, situation, etc. in a new way'. R11 also referred to "the acquiring of...behaviours", which further emphasised the reflection of C6. One other respondent (R17) noted that learning is "...a fairly permanent acquisition of something new, not just a *temporary change*", which is an element of C6. This same respondent, however, began their response with "My first instinct is to say '...'the acquisition of knowledge, skills or new ways to think', but I expect it's more complicated than that". Whether the respondent was anticipating the sort of answer that a researcher might be expecting was not clear but the development of the answer to the question indicated that the respondent was thinking through their answer and almost articulating their thoughts as they wrote as if they had not had the time or opportunity to think through this important issue before.

Skills was a term referenced in 15 of the respondents' answers to Question 1. Very few gave any indication of what these skills might be although one respondent did refer to '*higher order skills*' and some to the '*skills to use knowledge*'. In addition, there were 9 references to '*understanding*' in

relation to knowledge but no development of what 'understanding' means in this context. Assessment objective 1 in awarding body specifications across economics and business qualifications refers to 'knowledge and understanding' which implies that simple recall of knowledge is not the only requirement of learning but that some use of knowledge is expected although this is not stated nor developed in the specifications. Revisions of Bloom's taxonomy (Bloom et al, 1956), distinguishes between the ability to recall from the ability to use information or knowledge in the lower order skills. However, in no instances of cases where reference to understanding was mentioned was there any associated link with changed behaviour. Where understanding or skills are mentioned but not necessarily related to changed behaviour, it can be argued that such references are articulated in terms of, and informed by, the assessment objectives. Teachers can base their learning judgements and understanding in the context of the examination system which informs their daily work. If the assessment objectives were defined differently would this influence teachers' conceptions of learning?

A common feature of respondents' answers was to link knowledge and understanding with the ability to apply knowledge. Across awarding bodies, application of knowledge is covered in Assessment Objective 2 (AO2). Respondents referred to 'selecting and applying knowledge [to different contexts]', 'applying...to a problem/situation', 'developing the skills of application...', 'doing something with [knowledge]' and 'being able to recall and apply knowledge'. In my experience as an examiner, the skill of application is something that presents challenges to students. It can be difficult for students to visualise economic and business contexts when they have not been exposed to them and a number of respondents to Question 5 mention the lack of opportunity to expose students to real business and economic contexts in their teaching as barriers to teaching and learning.

Being able to apply knowledge is an element of conception 4 (C4) where subject matter is related to other parts of the discipline and to the real world. Many of the respondents' answers where application is noted as part of their definition of learning, use language which is very similar to that used by awarding organisations in descriptions of the assessment objectives for qualifications so it is not surprising that application should be mentioned in the context of any definition of learning given the closeness of teachers to the examination process.

Of the respondents whose answers show some reflection of conceptions 4 - 6, most are of C4 and relate to the skill of application as noted above. Other references identify 'making sense' of material (although in one case this was specifically related to making sense of case study material which can be interpreted as conception 3 (C3), which is a requirement of being able to answer questions appropriately. Relating 'materials to their environment and extend beyond their own experiences', 'gaining an understanding of the world around them', 'the interpretation of...knowledge and skills', 'what students can do that they could not do before' and 'where students find meaning to a topic in a positive environment' were comments that reflect C5.

The majority of respondent answers, as noted, tended to reflect a conception of learning that focused on knowledge acquisition. There were a number of comments added by respondents which reflected a conception of learning which was peculiar to the classroom rather than more generic. R1, for example, noted that learning can be accomplished by "...study, practice or by being taught", which implies that the way humans learn happens in a very controlled environment and not through experience, discovery or through the making of mistakes. Recall of knowledge was a common feature in respondent answers often linked to the ability to apply the knowledge. R2, for example, referred to assimilation of knowledge alongside the ability to "...select and apply that knowledge at the appropriate moment"; R3, R10 and R14 all referred to "using knowledge", R21 and R22 both referred to the ability to recall knowledge and apply it. R21, however, offered a definition of learning which differentiated learning in the classroom and elsewhere. Initially the answer explained learning as a situation "When we experience something,
understand it and then know about it" but then followed this with: "In the classroom, learning is being able to recall and apply knowledge to a particular situation".

6.2.2 Question 2

Question 2 was designed to investigate the extent to which this sample agreed on the key economic and business concepts which form the basis of any teaching and learning programme in the discipline. It can be hypothesised that there ought to be some common agreement amongst teachers of key concepts that form the basis of a discipline and which need to be covered in any teaching and learning programme in order to help students understand the subject.

The range of 'concepts' identified was wide with 53 items stated. Some of the 53 were related and there were some overlaps but the variety of concepts identified suggests limited agreement amongst this sample on what constitutes important concepts to teach. I wrote the term 'concepts' in inverted commas at the beginning of this paragraph because there also seems to have been some confusion over what is meant by the term 'concept'. Only three of the concepts identified by respondents in this sample (elasticity, the margin and opportunity cost) are the same as the ten threshold concepts identified by the Embedding Threshold Concepts Project (Davies and Mangan, 2005).

Marketing or the marketing mix was the most cited concept, featuring 13 times. Opportunity cost, supply and demand and the external environment were cited on 7 occasions each, cash-flow, profit and loss, finance and enterprise/entrepreneurship were each noted 5 times and elasticities and revenue, cost and profit were cited on 3 occasions. Table 6.1 lists the range of 'concepts' identified.

Table 6.1 Concepts identified by the sample

Concept	Number of
	references
Marketing/the Marketing Mix	13
Opportunity cost	7
Supply and Demand	7
*External influences	7
Cash-flow	5
Enterprise/entrepreneurship	5
Profit and loss	5
*Finance	5
Elasticities	3
Revenue, costs and profit	3
Competition	2
Added value	2
*Change	2
*Motivation	2
*People in Business	2
*Production methods	2
Stakeholders	2
Free markets	2
*Break-even analysis	2
*Business objectives	2
*Human resources	2
*Aims and objectives	1

*Macroeconomics	1
*Customers	1
*The basics of business (what businesses do)	1
The margin	1
Equilibrium	1
Incentives	1
*How to start a business	1
The difference between cash and profit	1
*Recruitment	1
*Operations Management	1
*The role of government and the European Union	1
Competitive advantage	1
*Economic understanding	1
*Solving business problems	1
*Types of business and business practices	1
*Choice and the process of production	1
*The Economic problem	1
*Planned economies	1
*Market economies	1
*Mixed economies	1
*Legal structures	1
*Strategic decision making	1
Risk versus reward	1
*Measuring performance	1
*International trade	1
Exchange rates	1

Interest rates	1
*Government policy	1
*Meeting customer needs	1
*Ownership	1
Costs versus benefits	1

Some of the concepts listed do have some overlap, for example, cash flow, profit and loss, costs, revenue and profit and the difference between cash and profit can all be seen to be related content areas. It can also be questioned how many of the list in Table 6.1 could be classed as concepts. Many (32 of the 53 'concepts' listed (60%)), could be argued as being content areas rather than concepts and those that come under this category I have indicated with an asterisk. If there is such a disagreement about what constitutes the core concepts of a subject and indeed what a concept is, then the learning outcomes on which teachers base their planning is likely to differ and in turn the purpose of assessment (the process of gathering, interpreting and using evidence to make judgements about students' achievement in education, Harlen, 2007).

6.2.3 Question 3

20 of the 22 respondents stated that they thought it important for students to learn key terms to tackle the examination effectively. The language used in respondents' answers noted that they believed it 'very important', 'the most important element', 'extremely important', 'absolutely essential', 'crucially important' and 'crucial!'. Some commented that learning key terms was a 'foundation' which enabled students to access the examination and without such learning students would 'struggle to access questions'. It was evident that some respondents suggested that the nature of the assessment and the examination made it important that students learnt key terms because it gave students 'confidence to complete the paper' and that it was required 'before [students] can develop higher order skills' and provides the opportunity to 'support business knowledge and give students a key to hang theory on'. This latter comment implies that key terms are seen as different from business knowledge if they are acting as a support to knowledge.

Focusing on key terms was a feature of the teaching programme for a number of respondents. R3, for example, noted that it "occupies much of staff ingenuity and skill" and that teaching key terms was "...embedded in the teaching/every lesson/activity" although R3 also noted that this must not become "...a chore". Other respondents said that they used tests in order to encourage students to learn key terms. R16, for example, stated that they "...give students key terms tests every week in the last 2 - 3 months before external exams in order to impress on them the value of knowing them" and R18 said: "I often use 10 definitions test at the start of each week from GCSE Business right up to A2 Economics".

Given the affirmative answer to the question set, it is not surprising that the answers reflected conceptions 1 - 3 to a large extent. Of the 20 respondents who said that learning key terms was important, 18 were noted as reflective of C3. In addition, there were elements of C2 and C1. Examples of these include R9 who stated that "...marks awarded for definitions account for over a grades worth (sic) of marks". RI2 said: "I encourage all of my students to memorise definitions so that they can gain the knowledge marks straight away, as I think this is the easiest part of the exam and can help them gain a high grade even if they are unable to analyse or evaluate to the required standard"; R13 noted: "Improving the quality of terminology usage has been instrumental in improving exam performance in my classes", R17 commented: "...it communicates to the examiner that the student 'knows what they're talking about' and therefore promotes the awarding of higher marks"; R18 suggested that they "...have my students learn each set of ten [definitions] by rote memory - as they did spellings and tables at primary school" and went on to say that "...definitions need to be so precise and clinical; textbook perfect - I find this an effective way to help boost examination grades".

7 of the 20 who considered learning key terms was important, gualified their answer. R1 noted: "...learning key terms in themselves will not necessarily enable a student to do well in the examinations. Students need to be able to understand the concepts as well as just the key terms. The ability to analyse and evaluate effectively ensures students will achieve the higher order marks". Here, the phrase 'doing well' in the examinations was a contrast to R9 above, who noted that learning key terms could make a difference of one grade in the examination and highlights different interpretations of teachers about what is meant by 'success in examinations'. R1 also makes a distinction between understanding concepts and key terms. There appears to be some difference of opinion on the extent to which key terms and concepts are one and the same thing. Some imply that the two are separate as suggested by R1's answer. R21 equates the two by saying: "If they can understand concepts they can apply them to a variety of situations and questions". R6 noted: "...repeating the definitions word for word is very limited in its usefulness; it is the deep understanding of the concepts that is the crucial thing". This comment not only justifies the need to understand but also implies that key terms and concepts are the same. R21 suggested that key terms were not just to be seen as "business terminology" but also the command words used in

exam questions. The respondent emphasised the importance of students' understanding these 'key terms' in being able to tackle the assessment.

R10 suggested that: "Key terms give a structure to the concepts the students are learning. It is very easy for students to use the word money in place of profit, costs, price, expenses to name but a few. If students understand the correct terms for concepts they are less likely to get confused and more likely to be precise about what they are talking about - e.g. cost when they mean price - these are very different terms". In this answer, there appears to be some distinction between key terms and concepts but an appreciation that they are linked. The phrase "if students understand the correct terms for concepts..." relates a precision in the use of language with the related concept. R10 highlights 'price' and 'cost' as being two 'terms' but it can be argued that both are key concepts. R19 recognises the importance of learning key terms but notes that this skill is not as important as "...the ability to use theoretical concepts in a given context" and then goes on to expand on this by saying: "By which I mean, recalling a specific word is not as important as being able to analyse and evaluate a complex situtation (sic)". This separates out the learning of key terms with other skills (using theoretical concepts in a given context) which is then clarified as the skills of analysis and evaluation.

Of the 7 answers where some expansion was given in justification, most referred to the need to 'understand' as well as 'learn' the terms. R21 is typical of this in noting that it is important students get "...the opportunity in lesson time to talk about the terms, discuss the meaning, engage with the terms and to then have the chance to "prove" their understanding by applying them". R20 said that it was "extremely important" to learn key terms "...but they need not just to learn but understand them". This answer seems to imply that learning and understanding are two different things. R5 linked understanding of subject specific terms with being able to 'access the examination'. R7 noted that knowing key terms by rote does not necessarily mean there is understanding and then suggests that this is important in relation to being able to answer questions appropriately. R11 acknowledges the importance of learning key terms because of the way the assessment is formulated. The respondent supports this by quoting the proportion of marks awarded by some awarding bodies for explaining key terms (20%) and so concludes that if there is a reasonable proportion of marks awarded for demonstrating this skill, then it makes sense for teachers to place a reasonable focus on it.

The two respondents who did not think it was important both made a clear link to the necessity of understanding or explanation rather than simply memorising. R2 said that explaining "...the idea behind the key term" was the important thing because "Marks are often awarded for variations of the key terms", which suggested C3. The other respondent, R14, placed their answer in terms of how other teachers think by saying learning key terms was "Not as important as most teachers/students think". The respondent went on to say that developing learning skills was of greater importance and exemplified this by suggesting: "High level students will succeed because they can adapt and transfer skills to solve problems and form opinions." which is reflective of C5.

6.2.4 Question 4

Every respondent said that they would find a way to use the student's question and article in their teaching in some way. R8 noted that "...a good *teacher would not turn down this opportunity*". Three said that they would use it at some other time than the immediate lesson. The vast majority, however, made reference to using the question and article in a 'starter' or '*plenary*'. The use of these terms indicated that the respondents have some structure to their lesson planning which includes this terminology. The Times Educational Supplement (TES) has a template document which is entitled 'The Perfect (OFSTED) lesson plan' (TES, 2013), which provides a guide to teachers for planning their lessons and taking into account key features in Ofsted inspections. The plan includes time allocated for a 'starter' and a 'plenary'. This seems to imply that any 'good' lesson must include these features and the frequency with which respondents mentioned starters and plenaries in their answers to Question 4 suggest this is a typical feature of most lessons of this sample. A report on Key Stage 3 by Ofsted (2002), states that it recommends a 'three-part lesson structure (a starter activity, a main activity and a concluding plenary)' and the responses to this question suggest respondents have this structure firmly in mind when thinking about lesson planning.

Nine respondents indicated they would use the question and article as a starter. In many cases this is linked to the intention to use the information as a means of connecting learning. The question indicated that the article and question related to a topic area that had been covered in a previous lesson 'two weeks ago' and that the teacher had now moved on to a different topic. Finding ways of using the question as a starter to help recap or connect areas

of learning was a common response. R2 said they would "use it as a starter to recap content from the previous lessons. Probably wouldn't spend more than 5-10 minutes on the discussion through (sic)", R6 noted that they would "...address it quickly at the start or end of the lesson", R8 said they would use it "...as a starter to remind students of the concepts learnt", and R10 answered that they would "Use it as an opportunity to recap earlier work assuming that other students were also interested and it provoked discussion in class". The caveat of whether 'other students were also interested' and whether the question and article would 'provoke discussion' is an interesting response given that it might be argued that the teacher has some responsibility to make questions/topics interesting. A later question on barriers to learning (Question 5) raises the importance respondents place on connecting real world events with their teaching and this question was designed to see the extent to which teachers are prepared to exploit a learning opportunity with their students against the plan that they have prepared for the lesson. R10, however, repeated that they would only use the issue raised by the student "...if students are interested".

The use of the article and question as a plenary suggests that respondents were prepared to bring the topic up for discussion at the end of a lesson. Ofsted (2002) notes that plenary sessions should be 'geared to giving pupils an opportunity to say what they have learned' which implies that the plenary should be linked to what learning has been achieved in the lesson itself rather than as an add-on to recap something new which a student has brought in. This might indicate that teachers in this sample have an incomplete understanding of the purpose and use of a plenary and indeed the same Ofsted report (2002), noted that 'plenary sessions remained the weakest element' of lessons observed in the pilot that was the subject of the report.

R8 suggested that they would incorporate it into the plenary "...to encourage others to research similar articles", R14 would "...use time during the lesson to think of a stimulating plenary to the lesson based on the article" which further implies a confusion between the purpose of a plenary as defined by

Ofsted and activities which constitute the end of a lesson but are not a plenary. R19 would "...answer the question to the individual at the end of the lesson" even though this same respondent expressed a very developed definition of learning in Question 1 where they said: "I think of learning as a process of brain building! Adding new neural connections to existing experiences; helping a student change way they process information; providing new strategies for decision making....I also believe that to learn, but not to do, is not to learn". Leaving a learning opportunity to the end of a lesson and then just questioning the student seems incongruent with this definition.

R1 explained that bringing in newspaper articles was an expected part of the teaching and learning programme at their school. However, this practice is at a set time each week (Monday morning) and so even if the student brought in the question and article at a different time of the week, the respondent said that they would postpone following it up until the following Monday, which implies a rigid approach to lesson planning and responding to student queries/interest. R8 also implied that the response would be 'postponed' in noting "If the student is particularly keen and cannot wait for the discussion *lesson...*", which suggests that particular time is set aside for discussion and that such discussion is not always seen as being a spontaneous thing. R3 noted that it is 'not uncommon' for students to 'bring issues from part-time work' whereas R18 said that such an incidence happens "...but not as often as I would like", and R19 implied such an occurrence was very rare by noting: "After I have recovered from my faint...'. Some respondents noted that they used display boards in their room to show relevant news articles and also to encourage students to bring in articles of relevance. R9 said they would pin up the student's article on their display board, R17 would "...use the article to add to the 'current affairs' display wall" and R1 explained that the articles students bring in are "...pinned on a display board for the current week".

A number of respondents said that they would celebrate the student's

initiative; R3 would "...praise them and thank them", R11 explained that they would share the student's initiative with the whole class as an example of good practice ("Spontaneous and unprompted inquisitiveness is to be celebrated! This should be used to model the type of approach that is desirable"), R17 would 'praise and thank' the student and R18 would "...reward such enthusiasm within the school effort and attainment grading system". However, only 2 respondents said that they would completely abandon their planned lesson and focus on the student's article. R4 suggested they: "Would definitely abandon my plans and do the lesson I planned and focus around the news article - maybe not all the lesson but long enough to do it justice" (perhaps a confused answer) and R11 said they would "Hold fire with what had been planned". In most cases, therefore, the student's query would be built into the planned lesson in some way, most typically through a starter or plenary.

6.2.5 Question 5

Looking at barriers to improved understanding provides an opportunity to investigate whether there are common features which teachers see as being important. Some of the answers by respondents focused on external factors over which teachers have little control such as the nature of the specification and syllabus and to an extent the time allocated to teaching and learning. Some answers focussed on subject specific issues such as numerical skills and the amount of terminology required or the understanding of graphs. One particular feature, however, was the number of respondents who saw the barriers to improved understanding lying with students themselves rather than anything which would be in their own power to influence. In contrast, a small number of respondents acknowledged that teachers themselves can be the source of barriers to understanding.

In setting the questions, I anticipated time being a factor which might be a common theme but only three respondents, R1, R4 and R20 mentioned it. By far the most common barrier as perceived by the respondents was the lack of prior knowledge or experience. Some specifically noted lack of prior

knowledge as a factor whereas others expressed this as a 'lack of life experience/intellectual immaturity, lack of experience of the real world or an ability to relate to the real world what they learn in class'. In almost all the respondents' answers to this question, the focus was on the **student's** inability to relate learning to the real world. In some cases, such as R1, R3, R4, R12 and R20, the lack of space in the curriculum prior to Key Stage 4 for economics and business was identified as being problematic when trying to teach students at Key Stage 4; respondents noted that much of the material to be covered would be new to students. It was agreed that there was a lot of material to cover in a relatively short space of time.

This was linked to other comments that students have 'limited life experiences when they are aged 14 - 16' and this can mean they have 'difficulties linking what experiences they have had with what they learn in the classroom'. R6 said that this "...lack of intellectual maturity" was not something "...that could be addressed easily". R7 put the focus squarely on the "...inability of students to relate both Business and Economics to the real world around them", as did R16 who stated that a barrier was "...a lack of willingness/commitment on the part of the student to engage with all the real-world situations in which business and economic phenomena abound". R9 simply said "Lack of real life experiences", R13 linked a "Lack of awareness of 'real world' businesses" with an inability to "...apply their knowledge to a range of situations. This in turn can make their answers superficial", R15 as "The ability to actually view theory in practice in a business setting", R17 noted the "Limited knowledge of real world context", R21 as a "Lack of understanding of real world events" and R22 a "Lack of "worldly experience" and "Lack of opportunity to apply understanding to a real business".

What was not evident in any of the respondents' answers where this was cited as a barrier was any acknowledgement that in fact, students interact with economic and business events throughout their lives. The role of students as stakeholders in business was not a perception that the respondents in this sample provided. In addition, the responsibility that teachers might have to bring out the everyday experiences of students as stakeholders and relate these to the teaching and learning programme was also not acknowledged by the respondents in this sample. R22 did show some awareness that students interact with business because they "...go into shops or visit places such as a zoo..." but then suggested that students "...don't think of them as businesses". R3 identified the lack of prior learning as a barrier and added that "...it is all new and appears to be a novel way to work/think - why I have never understood but the students tell me so".

There was a desire expressed by some respondents to bridge the gap between the classroom and the real world by increasing the contact between schools and businesses. R1 'wished' that they were "...able to visit more businesses so that students can experience the different aspects of running a business", R15 saw a barrier as "The ability to actually view theory in practice in a business setting. Experiential learning and observing theory in practice would allow studnets (sic) to put the theory in to context and view the separate areas we teach as a whole interactive process" and R22 suggested that a "Lack of "worldly experience"" and "Lack of opportunity to apply understanding to a real business" could both be overcome "...by getting out into the world of business, either as a school trip, or using work experience as an opportunity to "experience" business, getting business speakers in, making teaching relevant to the world of business with real examples of real businesses (local if possible)".

Some respondents also 'blamed' students for their lack of ability to articulate meaning. R10, for example, commented on "...a student's ability to explain what they mean [as a main barrier]. They will often think a single statement is an answer to a question and will not reason through why they have come to a particular conclusion. They will give an answer to a question without including the steps in between which has got to that answer". R14 said that one barrier was students' "...ability to apply knowledge and contextualise an answer." and R18, the "Inability of students to transfer knowledge from

theory to application". R16 suggested that students have difficulty "...in 'scaffolding' from basic knowledge through to application/analysis, and then synthesis, and finally evaluation (including conclusion/overall judgement)". R21 noted that students have difficulty in being able to "...evaluate in depth. Getting them to think through consequences of decisions".

The implied blame on students as being the source of a major barrier to understanding could be seen as being a barrier in itself and one which teachers of economics and business need to tackle if student understanding is to be improved. It could be argued that part of the responsibility of a teacher is to help students make the connection between their real world experiences and the material covered in the classroom as part of a specification and to help students articulate their understanding. Seeing students as being the 'barrier' focuses attention on students' lack of ability rather than the teacher questioning their own skills in being able to help students overcome the perceived barriers.

In contrast, some respondents did reflect on the role of teachers as being a potential barrier. R2 noted that some teachers "...teach to the exam" which "...means that students have a narrow appreciation of what business studies is", R19 noted that an "Over focus on exam technique" was a barrier to understanding which could be interpreted as a problem with the system. R6 made reference to the quality of teaching and learning in business and economics as reported by Ofsted by stating: "Teaching and learning is often not good enough (from the OFSTED reports that I have read!)" and R17 said: "Honestly, my personal experience is that teachers can be a big barrier for students" and went on to explain their response by giving some examples of where poor teaching practice can lead to limited understanding in students.

Other common themes included the numerical skills required, (R2, R5, R8, R10 and R18), the amount of terminology that students needed to learn (R5, R8, R9 and R13) the nature of formal assessment which directed teachers to having to teach to the exam (R4), a lack of understanding of graphs (R8) and R12 noted that parents were a barrier to student understanding because they

"...think that business and economics is an easy option to take and so encourage their children to take it in order to do well". The respondent followed up by suggesting that this led to "...the wrong students taking the subject" which in turn "...effects(sic) their ability to understand concepts because they are expecting an 'easy' subject with very little work to do, when the opposite is actually true".

6.2.6 Question 6

'Thinking in the subject' is a term used by the UK qualifications regulator, Ofqual, in relation to a goal of learning. Thinking in the subject implies a transformational view of learning where students' views of the subject are changed to an extent that they will not view the world in the same way again and apply a series of processes to viewing issues and problems not just in the classroom but in their everyday lives.

16 of the 22 respondents specifically agreed that thinking in the subject was a key focus of their teaching and learning programme. Phrases such as 'very important', 'absolutely essential', 'crucial!', 'hugely important', 'essential', 'extremely important', 'vital' and 'vitally important' were used. However, subsequent expansion by many respondents suggested that their understanding of what thinking in the subject means varied and was at times confused. None of the respondents saw thinking in the subject as transformational learning although R17 did note "I think it's very important indeed- probably the core focus without which students will only ever be able to gain a surface understanding of the subject". However, the respondent then went on to expand their answer which suggested some confusion in understanding as they saw it as something which could be 'taught explicitly' and noted: "I'll definitely give this a really good think...and see how I might be more explicit about students stepping into different shoes in different classes, or whether this is necessary if the foundation of my teaching assumes this approach".

The idea of thinking in the subject as being a means by which students can put themselves into the position of an economist or business person pervaded a number of respondents' answers. R1, R4, R7, R9, R14, R15, R16, R17 and R22 all suggested that they saw thinking in the subject as a means by which students could take understanding and then transplant that to seeing perspectives from that of an economist or business person implying that these were people who were separate or external to what was going on in the classroom and who thought in a particular way which was not also relevant to the way students could be helped to think as a result of the teaching and learning programme. I am going to call this perspective *transplanted learning*.

Examples of transplanted learning include R1 who explained that their department was set up as a business where the respondent's view of what happens in business was established as the norm for the students. *"This means the students are expected to be smartly dressed, with uniform worn correctly. Punctuality, attendance and the meeting of deadlines are expected to be adhered to"*. Different methods of teaching were described including the use of case studies and the use of role play to provide

"...practical opportunities to test business and economic theories and group or team work so that the ideas come alive and students can see them working in practice". The respondent finished the expansion of the answer by stating: "It is important to nurture the entrepreneurs of the future, to give them the necessary tools they need to think about starting and setting up a business for themselves", which implies that the respondent sees the purpose of teaching economics and business as developing future entrepreneurs and business people rather than developing a way of thinking. Developing future entrepreneurs implies giving students the information and skills needed to set up in business (probably not something that the vast majority of students will do when they leave full-time education) whereas encouraging thinking in the subject provides students with the skills and thought processes to approach problems and situations with a set of processes and tools which will help solve problems. One such problem could be 'how do I set myself up in business' but thinking in the subject is not limited to this goal. Other examples of transplanted learning were observed in R4 who noted that they tried to "set tasks as often as possible to but (sic) pupils in role where they need to think like a business person". R7 said that they felt it was important to "...look at every situation from the perspective of the business person/economist", R14 not only said that it was important to "...place the students 'in the shoes' of a business person and to gain a clear understanding of what the problems are." but then went on to say that this allowed students to "...give high level answers without any reference to specific knowledge/key terminology". R15 made a direct comparison between giving students the practical experience of running a business and how this would help students' understanding. "Motivation theory is much better understood along with the choice of method when they have acted as a director of a business" was the closing line to R15's answer to Question 6. It could be argued that very few teachers of economics and business have acted as directors of business but still teach motivation as a topic and it is highly unlikely that any 14 - 16-year old student would ever have the opportunity of being a director in a business save the somewhat artificial experience offered by programmes such as Young Enterprise. R16 said that thinking in the subject "...helps students to develop an innate understanding of what makes businesspeople and economists 'tick'" and R22 offered a developed explanation of how practical experiences of running business such as those offered by enterprise based programmes available to schools could help "...allow students to have a taste of what it is like to set up a business and learn what it involves.". They then went on to say that "Lots of students run activities like this without realising that it is like running a business (eq charity cake stalls, car wash, etc), so it is up to the school / teachers to make it more explicit". R22 continued by referring to TV programmes as a means of providing a useful insight into the world of business. Programmes such as 'The Hotel Inspector', 'Undercover Boss' and the "...more "obvious"" 'Apprentice' and 'Dragon's Den' were cited but it was not clearly acknowledged that these programmes are developed to be entertainment and of necessity tend to show the extremes or pick particular scenarios to

generate entertainment effect rather than be reflective of 'real' business.

Some respondents made a clear link between thinking in the subject and how this could improve exam grades (C3). R2 said that it was "Very important to achieve the best grades" and was "...an integral part of preparing students for the exam"; R9 made a specific reference to encouraging students to "think as a business consultant" (which is a very different concept to 'thinking in the subject' and another example of transplanted learning) and then explained that this was important so that students could "...make justified recommendations in their examination". R12 said that they "...try to encourage my students to think like an economist on a regular basis" which implies that their teaching and learning programme is not predicated on this process but is seen as a 'drop-in'; the respondent further suggests: "I tend to use the 'think like an economist' as an extension task for my more able students as it challenges them at an appropriate level and prepares them for their exam" and "If I had a more able group, I would like to encourage 'thinking in the subject' more". It is not made clear why there is a belief that 'able students' are more capable of thinking in the subject than 'lower ability' students. R16 thought it important as "...this aids rigour on the part of students in tackling questions and problems..." and R18 linked an explanation of their understanding of what thinking in the subject meant with the benefit to students' GCSE assessment.

The responses quoted above suggest that there are different interpretations between respondents in this sample of what thinking in the subject means. There is also some evidence that understanding is confused as to the meaning of the term. R8, for example differentiated between economics and business, seeing thinking in the subject as being "...less of an issue in Business studies where transmissions are less vital and memorised material is often examined" and noting that "...students who think like an economist value analysis and logical thinking over opinion" which seems to imply that thinking in the subject in business does not require the same skills. R13 said that they thought it was desirable but that "...subject knowledge is more important because without knowledge of the concepts then it's impossible to make appropriate decisions". R13 hinted at some frustration of being constrained by the nature of the assessment and turned the answer to a justification of process in relation to how they saw other people judging teachers' success: "...many of the GCSE questions are based on recall/explanation; I don't want to be a slave to the GCSE syllabus, but ultimately that's how my effectiveness as a teacher is judged". The respondent then went on to contradict themselves by suggesting that "...those pupils that are able to 'think like a business person' do produce the highest quality answers, so I try to factor it into my teaching". If students who think in the subject produce the highest quality answers and the respondent feels that they are judged by results then it would seem to be a logical link that making thinking in the subject a focus of the teaching and learning programme would improve results contrary to the respondent's initial answer that "...it is not the most important focus".

R18 made an analogy between thinking in the subject and speaking a language but seemed to divorce thinking and speaking. "I would suggest not just thinking in the subject is important, but the ability (sic) to speak the subject". The respondent explained that they felt learning words by rote preceded "...stringing words together [to] form a coherent French sentence that makes sense which helps them with their GCSE assessment". Throughout the answer was couched in terms of C3 where the abilities being described were all linked to performance in exams rather than a way of thinking. Having suggested that 'speaking the subject' was also a pre-requisite the respondent then suggested that "Students cannot gain good grades/marks [C3] unless they can 'speak' the Business and/or Economics language fluent (sic). In that, using key terms, using precise and clinical definitions, referring to reading outside of the core text, applying their knowledge to different Global *Economies or National Businesses..."*, which seems to imply that 'using precise and clinical definitions...' was an example of thinking in the subject. The respondent then shows further evidence of some confusion by noting that "I would agree the 'thinking' precedes the speaking - but to me they go hand Page | 235 *in hand*" - if one thing precedes another can they also go hand in hand? The respondent finished the answer by saying: "Pupils often think they can 'blag' it regarding language, but they are rarely awarded for it - regardless of how acceptable the understanding - unless the correct language is used and the correct frame work applied - it is not good enough", which implies that if an answer to a question in an exam shows understanding the student is not rewarded for the understanding if the correct terminology has not been used. It might be asked whether an answer which shows understanding could be given without using any specific terminology appropriate to the subject area?

Overall, therefore, there appeared to be a number of interpretations of what 'thinking in the subject' means and little by way of agreement by this sample on a way in which a discipline is held together, its methodology, its way of approaching problems and questions and how this might help to benefit students in tackling assessment.

6.2.7 Question 7

A greater degree of agreement from respondents about features which impact on their teaching was evident from the responses to Question 7. Five respondents mentioned time as being the one change which would improve students' learning. 2 - 2 ½ hours per week teaching time for GCSE is a standard time allowance although this does vary across schools. R2, for example, lamented the lack of time on 2 ½ hours but R1 only has 2 hours of time and noted that even " an extra hour, if only for one year" would make a difference for them and allow more time for "...more in-depth investigation and research to take place as well as being able to revisit learning".

R1 went on to explain that more time would allow opportunities for visits to businesses and trips out. 4 respondents noted that opportunities to visit businesses, arrange trips or have businesses/people coming into school would improve students' learning. As noted earlier, it seems that respondents in this sample divorce students' everyday experiences and their daily interaction with economics and business from what goes on in the classroom and it may be that creative thinking in the way in which the real-life experiences of students can be harnessed to help drive improvements in learning do not appear obvious to some teachers in this sample. How far this is generalizable across other economics and business teachers is unclear but an Ofsted report in 2008 on the quality of teaching in the disciplines noted that despite the general enthusiasm of students for the subject, and whilst a majority of the teaching observed in the survey was good, "too much of the rest was uninspiring" (Ofsted, 2008, p4).

The other main theme in this question is the lack of technology and specifically, access to computers cited by 3 respondents. It might be suggested that the access to technology for many students via their smartphones is a possible option to overcome this particular issue but schools do not always see the potential of this technology as a learning tool and instead seek to ban or control its use, assuming that the technology will not be used for positive learning related to the subject being taught.

Other responses to this question generated a range of different problems and issues. R6 complained that they did not have sufficient information related to controlled assessment. "Exam board policy seems to be to let pupils loose with a completely free choice of question with very limited guidance. Reconciling this with doing your best for the pupils whose parents are paying £13,500 a year is, frankly, impossible. The c/a [controlled assessment] marks of my pupils are consistently below where they should be." This response is interesting because the amount of feedback, training and guidance made available by the awarding body concerned following the introduction of controlled assessment was extensive. In addition, the reason why controlled assessment was introduced was because coursework, which it replaced, had its validity and reliability as an assessment called into question because of the extensive support given to students by teachers. There seems to be a misunderstanding by this respondent about the nature and purpose of controlled assessment and the policy imperative for introducing it. Instead, the respondent blames the marks students are getting on the awarding body rather than accepting that teachers have a responsibility to understand the

assessment and support students appropriately. R10 suggested reintroducing coursework "...so that students investigate a business problem in the real world". There does not seem to be recognition by the respondent that there is nothing preventing any teacher from investigating business problems in the real world as part of any normal teaching and learning programme linked to an examination course and that for some awarding bodies the investigation of real world business is fundamental to controlled assessment.

R12 suggested that the ability of teachers to be selective in who was allowed to do the course would improve the learning of students given the challenges facing teachers with mixed ability classes. This respondent was clearly focused on the importance of getting appropriate grades for the students. R16 said that a requirement for students to "...keep a tidy file of news reports and other relevant articles, which they regularly added to as their courses of study progressed" would help improve student learning. It was not clear how such a requirement would improve learning albeit that the respondent added that: "...files would be reviewed periodically by myself, with detailed feedback being given to the relevant form tutors, and summary feedback being given to the Assistant Head (Academic)". Where the students stood in this suggestion was not clear.

Other respondents noted the problems faced by policy issues - whole school intervention policies at Key Stage 4 (R19), the segmentation of Key Stage 3 and 4 (R14) and a *"…less crowded curriculum"* (R11). These are external factors over which teachers feel they have little control but which clearly seem to affect the respondents' view about how student learning could be improved.

6.2.8 Question 8

19 of the 22 respondents saw syllabus coverage as being 'very important' with some suggesting it was 'crucial' and 'extremely important'. Many of the responses to this question reflected C3 with 20 of the 22 responses being tagged as having some resonance with this conception. Respondents noted that coverage of the examination syllabus was important for giving students confidence in being able to tackle the examination successfully and in some cases, to help avoid students panicking when coming across material which is unfamiliar (R9 and R13). R1 noted that feedback from 'personal research' with students had been that students themselves felt that: "...it is important to be fully equipped to do as well as possible in their examinations by having the necessary subject knowledge as well as having had the experience of practice papers and questions and the correct examination technique" and R3 explained that support is given to students to organise their material to aid revision and improve chances in the exam. R3 also noted that they "Would get huge complaints from parents (rightly) if we had not covered all of syllabus".

R4 noted that it was a "...focus of my planning, sometimes to the detriment of quality learning" and R14 also acknowledged a self-contradiction in saying: "It is worrying in that there is the constant fear of 'what if there is a question on it'? The traditionalist in me still places a lot of emphasis on this despite some comments earlier which focus on developing learning skills rather than covering specific content." R18 said that it was not of great importance for them as a teacher but was in ensuring pupil and parent confidence.

6.2.9 Question 9

Question 9 was written to deliberately test the extent to which teachers in this sample reflected C3. Writing frames assume that there is a set approach to writing answers to questions, in particular those which require higher order skills to be demonstrated. The structure of writing frames tends to reflect the schema referred to in Phase 1 of the study. 20 respondents said that they used writing frames in some form or another. Respondents noted that writing frames help build confidence, help students to write in detail, provided the student with key phrases and terms to use, prevented the student from going off-task and were useful in helping students write extended answers where marshalling thoughts and organisation were important elements of an answer. Of the 2 who said they did not use frames, R9 noted that they used a variety of structures and allowed students to choose one that suits, which suggests that some form of writing fame is used. R22 also said '*no*' but then went on to explain that their predecessors had used them and that they believed there was some value in using writing frames for coursework rather than exams but then somewhat contradicted the answer by suggesting that writing frames were not so useful in controlled assessment (a variant of internal assessment of which coursework is a type) because they led to a "...formulaic response by the students, i.e. a lack of individuality". The fact that coursework and controlled assessment are both forms of 'exam' seemed not to be considered by this respondent, the implication being that internal assessment.

Of the 20 who said 'yes', the vast majority referred to the value of writing frames in providing a structure or scaffolding to help students and a number suggested that writing frames were more useful for weaker students. R12 made a clear distinction between the value for weaker students in providing "...guidance and structure" and the restrictive nature of writing frames in relation to more able students who may not be able to "...discover things for themselves". R14 felt that writing frames were particularly useful for boys "...who are reluctant to add detail to extended writing responses".

R7 mentioned that there were benefits to the teacher of writing frames as it was "...easier for the teacher to see if the student has understood, because they are easier to check than a rambling piece of work with no structure". R18 said that they thought writing frames were suited to a "...more didactic teaching style" and commented that they tended to be too passive and restrictive and made for 'uninteresting lessons' but then went on to say that they were suitable for weak students who, it might be suggested, are the students most in need of active learning approaches and lessons which are interesting.

Despite the fact that most respondents used writing frames and offered explanations which justified their value, there were a number of cases where the respondent offered some balance to their answer by noting the potential disadvantages of writing frames. Most typically, the main disadvantage seemed to be the feeling that writing frames could become a 'crutch' which students could come to rely upon and not develop independence as a result. R3 noted a variety of structures that were in use including providing students with cues, providing resources which helped students prioritise and the provision of "...key analysis and evaluation phrases/questions on all classroom walls which we use as prompts and direct students to every lesson" but then went on to pose an interesting question: "Are these writing frames? Or logical thinking translated into writing?". R7 noted that "...they should be used judiciously and sparingly, only when the teacher feels that the concept/subject area is particularly challenging. It should be explained to the student why the frame is being used and they should be constructed with great care to fit each differing situation". R8 mentioned that "...they do need to be used sparingly as students can become overly reliant". R13 said: "I think it's a helpful scaffold, but only as part of a long-term process that aims to move pupils towards complete independence". 9 respondents offered some additional explanation in the form of a caveat which recognised the value of writing frames in some instances but were cognisant of the limitations of such methods.

6.2.10 Question 10

10 of the 22 respondents commented that they had no issues with the question types that appeared on exam papers. R1, for example, said: *"There is a good variety of questions, with real life case studies which range in difficulty and I feel is a fair test of a student's ability"*. R8 and R12 did add that they would like to see less emphasis on multiple-choice questions (MCQs). R8 said that they believed these types of questions *"can be very challenging and confusing for students"* but R12 noted: *"Although* [MCQs] *provide students with a higher chance of gaining some marks, they are not stretching or challenging enough in my opinion"*. MCQs are used in GCSE examinations as a means of primarily assessing knowledge and understanding, and some application (AOs 1 and 2) and provide the means by which

examiners can target a range of ability levels; MCQs with relatively simple distracters and key can be useful in allowing weaker candidates to access the assessment but can also be used to discriminate effectively between different abilities when questions are constructed carefully. They can, therefore, be challenging and stretching.

The remaining respondents offered a variety of suggestions for different types of questions that they would like to see included. R2 wanted to see "*Open ended strategic questions*" and qualified this by adding: "...to my mind think strategically is a vital component in business". R3 and R11 said that they would like to see more questions where data was presented and required some interpretation by candidates. R3 commented: "In all papers would like to see opportunity for candidate to develop own solutions more rather than evaluate options from a case study" and R4 wanted to see "Questions which develop their entrepreneurial skills - coming up with ideas for new products etc". Some respondents referred to the need to use more 'real' information in questions; R9 wanted to see the "...use of more real businesses" and R20 wanted: "More realistic questions regarding topics such as recession, redundancies, and current economic climate".

Several respondents noted that the numeric and mathematical skills necessary for studying and understanding business and economics were not as prominent in questions as they could be. R3 wanted "...more interpretation of data/number - it is an essential skill in our subject and too often when it is assessed it is via accounts" and R14 said: "I do feel there should be more financial questions which rely on construction of financial statements rather than interpretation. Interpretation answers tend to be taught in a more mechanical way whereas the request to construct the statements requires a greater analytical understanding of finances, particularly when problems occur and need to be rectified". R18 was particularly critical of the lack of mathematical and numerical assessment in exam questions stating: "More mathematical elements to the Business and Economic course - there seems to be much focus on writing ability with less emphasis on numerical ability - business is about the numbers, the profit, the bottom line. This, in my view seems to be missing. Economics likewise, every undergraduate degree in economics is heavily reliant on maths - but GCSE...it contains so few numbers - it is the subject diluted down so it's almost unrecognisable. To even call it economics at GCSE level (a social science where we study the allocation of scare resources) is laughable". Such views provide an interesting contrast with answers to Question 5 where respondents noted numerical and mathematical skills as a barrier to learning, indeed, R18 responded to that particular question by citing "Fear of numbers" as being one of the barriers to learning.

Other suggestions for additional questions included the use of more extended answer questions (R12 - "I believe that there should be more 8-10 mark questions requiring the higher order thinking skills in order to distinguish between good students and excellent students"), R13 commented that they thought the range of questions available were appropriate but that "...the only significant omission is long essay questions, but I am not sure these are appropriate for an exam paper with no tiering (I suspect essays would be a great discriminator between those who are A and A^{*} grades, but are totally inappropriate for those who are aiming for a D or E grade)". R16 wanted to see "...a separate 'Critical Analysis and Evaluation' paper, where the nature of the questions set challenge students to not only carefully analyse the information given, but also make careful judgements, taking into account the weaknesses in the information present (for example, data presented has missing years or is in other ways limited, partiality with views expressed, 'biased' sources, and so on)". R17 said they thought the range of questions was adequate but that "I've noticed that graphs / charts are getting tougher, and calculations have more stages now. I think that this definitely tests understanding more rigorously so I support this", R19 suggested some supported multiple choice questions would be useful: "...as a way of quickly testing ability to apply a range of theory", R21 wanted "More that encourage understanding of diagrams and current economic issues" and R22 suggested that: "pre-release material would be a useful addition so that "students do Page | 243 not have to "pull their understanding out of the bag" on the day of the exam; weaker students find it difficult to read lots of text in a short amount of time and then apply their understanding in any depth. This discriminates against these weaker students who are penalised for both their weak literacy and weak time management skills".

Whilst 45% of respondents felt that the range of questions available was adequate, there did appear to be some differences in the views of respondents on what sort of additional question types would be useful and the reasons why the question types suggested would be valuable. Some of the reasoning given in proposing different question types highlighted a lack of understanding of the rationale for current assessments and the way in which exam papers are put together. Examiners have to recognise that for a paper to be valid and reliable, it is likely that the marks gained by the cohort sitting the assessment will follow the typical bell curve that comes with a normal distribution. Not every candidate will be able to access every question and it is not desirable that every question be accessible to all candidates - the exam would not do its job of discriminating if that were the case. The range of question types in existing assessments also has to reflect the balance of the assessment objectives specified in the criteria laid out by the regulator. If more extended writing questions were included at the expense of, for example, MCQs, the question setter would have to ensure that the assessment of knowledge and understanding (AO1) that is a feature of the use of MCQs could be adequately addressed in these types of questions. More emphasis on numerical and mathematical skills would have to be more explicitly stated in the specifications and given the subject criteria in existence at the time of writing, examiners are not in a position to provide questions which assess these skills as they are not explicitly required.

6.2.11 Question 11

Question 11 was designed to tie in with Question 1. Question 11 asks respondents to explain how they recognise when learning has taken place in their lessons. In order to answer this question, logic would suggest that having a clear understanding of what learning means to the individual would be a prerequisite for observing and identifying when it occurs in students. Part of the analysis of this question also looks at the answer given by respondents to Question 1, which asked them to explain what they understand learning to be, and judging the extent to which the answer to Question 11 is consistent with that given for Question 1.

Common responses to the question included the use of questioning techniques as a means of identifying learning both relating to questions asked of students and students asking questions. Confidence and the ability to articulate through either verbal or written means were also noted by a number of respondents and several also mentioned peer assistance where students explain concepts, issues or problems to their fellow students. Many of the respondents' answers included some reference to being able to tackle written tasks, exam questions or perform better on tests as being evidence of learning; these responses were suggestive of C3.

There was little indication that respondents saw a change of behaviour as evidence of learning, reflecting conceptions 4 to 6. R3 presented a range of situations where learning might be witnessed and noted that it did depend on the student. The different situations included:

- 1. Willingness to look up and offer answers
- 2. Developing own analysis and evaluation either orally/group work or in writing
- 3. Using technical vocabulary naturally and accurately (sic)
- 4. Helping others
- 5. Quality of written work
- 6. Asking/answering the next question without prompting
- Asking is this right or wrong but being prepared to try independently first
- 8. Applying materials to different situations

- 9. Asking questions such as "what if"
- 10. Bringing own examples into discussions
- 11. Test results

Points 1, 5 and 11 could be argued to reflect C3, point 2 C4, point 3, C1, points 6, 7, 8 and 9 could be said to be reflective of C4 and C5. This respondent's response to Question 1 was largely reflective of the mix of answers to this question; R3 noted in their response to Question 1: *"Student develops the skills to use that knowledge and/or to make sense of case study material to apply it to a problem/situation and be able to make logical and informed comments/analysis or evaluation"*, which could be indicative of C4 but the reference to the specifics of making sense of case studies and the link to higher order skills specified in the assessment objectives is still reflective of C3. However, R3 then provides some further detail which is reflective of C5 by noting: *"Student can relate materials to their environment and extend beyond their own experiences"*. R10 stated: *"The knowledge and skills are transferable to other situations and problems"* which is reflective of C4.

R13's answers to both Question 1 and Question 11 also showed some evidence of a match. For Question 11, R13 wrote: "The million dollar question! For me, it's if they can take a concept and use it (rather than repeat it parrot fashion) e.g. if they can use the terminology correctly in a different context, or use a formula to calculate something, or explain a concept in their own words" and for Question 1: "When a pupil develops a better understanding of a concept, masters a new skill or improves an existing skill." The focus in each answer is on concepts, using concepts in different contexts (C4) and developing a better understanding of concepts.

R21's response to both questions showed some divergence. In response to Question 1, R21 made a point of referring to learning in two different contexts by explicitly distinguishing between 'learning' in general and 'learning in the classroom': "When we experience something, understand it and then know about it. It could be an action or an experience or knowledge. In the classroom, learning is being able to recall and apply knowledge to a Page | 246 particular situation". In response to Question 11, R21 wrote: "Progress. When they get it. When they understand and can explain something they have struggled with. When they complete an assignment to a high standard. When they ask good questions and demonstrate a thirst for knowledge". In this answer, the reference to understanding and explaining, is made in relation to classroom learning and the focus on completing assignments is related to C3. However, the reference to "When they get it" and "...demonstrate a thirst for knowledge" are more reflective of C4 and C5. R22 also noted: "When they are able to transfer their learning from one context to another. When they are able to make links between what they've learned in a lesson with what we've learned in previous lessons", which is reflective of C4.

Given that the explicit context of this question is in relation to learning in the classroom it might not be surprising that a more general view of learning in relation to broader conceptions of learning articulated by C4 - C6 are not prevalent. Many of the answers focus on the typical activities one might expect to see in the classroom such as students asking and responding to questions, their written work, performance in tests, how confident they appear to be both in their ability and willingness to respond to questions and how they work with groups and with their peers. All these responses tend to reflect a conception of learning which is related to conceptions C1 - C3. Some respondents referred to the use of starters and plenaries as ways in which they observe whether learning has taken place. R1 noted that: "The starter can be a recap of previous learning" and R15 said: "Recap and structured questioning gives a quick check of students understanding", the implication being that in using a series of questions at the start or end of a lesson, this provides evidence of learning but in a broader definition of learning which relates to changed behaviour, it is doubtful whether simple recall of knowledge in response to what are likely to be relatively closed questions given the time available and the number of students typically involved, is indicative of a change in thinking.

In other answers, the use of questioning as evidence of learning has been

noted earlier and the type of questions used in the classroom can be important - the balance between closed and open questions used, for example. Peer assistance, participation in discussion, being confident in presenting material and ideas were all mentioned by a number of respondents as evidence of learning. However, the use of observation and listening was only mentioned by one respondent. R15 said: *"Group work allows me to listen to discussion and assess application of knowledge"*.

Only 3 respondents specifically mentioned formative assessment and in two of these responses there could be some doubt cast as to whether formative assessment was fully understood. Only R11 seemed to be clear on the meaning and use of formative assessment in stating: "These [ongoing formative assessment] opportunities might come through questioning, review exercises, discussions. Students must be provided with an opportunity in every lesson to "demonstrate" their learning ... Information from this phase of the lesson will tell the teacher whether learning has taken place, and will inform planning of subsequent lessons". In the other two cases where formative assessment was referred to, respondents mentioned the use of written work as being reflective of formative assessment. R1 said: "More formative assessment in the form of tests could also demonstrate learning" where it could be argued here that tests are more reflective of summative assessment than formative. Similarly, R14 mentioned "...formative assessment of written work". In relation to the specific question being asked in Question 11, it is not clear that in these responses, formative assessment is being seen as a means of planning and providing feedback for the teacher but instead is being seen as an indicator of what learning has taken place (summative assessment). Only one respondent mentioned assessment for learning (AfL) explicitly. R2 referred to "AfL mechanisms" but gave no further expansion on the point.

6.2.12 Question 12

The question was framed to refer to an individual student but a number of respondents included how they would deal with groups of students in their

answer. Whether this was a case of respondents not reading the question carefully enough or reflecting on the more typical way in which they deal with issues where confusion with economic/business concepts arise is not clear. Initially, I will focus on the response given in relation to dealing with an individual student.

12 of the respondents said that they would give written feedback which by implication suggests they have assumed that the confusion comes to light as a result of a written submission by the student rather than coming to light through other formative means such as in classroom discussion and debate. In some cases, respondents then noted that they would provide further written work to assess whether the confusion had been tackled. R7, for example, said: "...I would provide them with some more written work, possibly constructing a framework to structure their writing. This would test their understanding of the concept and hopefully provide them with confidence that they now understand the concept". The idea of providing some sort of model or exemplar answer as the means of dealing with the confusion was also mentioned by R9 and R16, and R21 noted that ultimately they would "...tell the student the answer". The question did not refer to giving a correct answer but was focused on confusion over a key economic/business concept. In these cases, it can be argued that the teachers are demonstrating C3, where acquiring methods of answering questions is at the core of this conception.

A common way of dealing with the issue was to have some sort of one-to-one meeting with the student to talk to them and try and solve the confusion. This was not always the first response to the problem but one which is used when other tactics have been tried such as written feedback. 13 respondents mentioned this type of tactic with lunch time and after school being popular times when such meetings would be arranged. R17 said that they would find the time to interact with the individual during the lesson by planning "...with some independent work time so that I could have the rest of the class working on that while I sat with the one student to re-teach the topic". Most

comments relating to one-to-one meetings involved discussions with the student to ascertain the nature of the confusion and then a variety of tactics to help assess the extent to which the confusion has been overcome. Such tactics include setting further written work, using the confusion as a means of covering the topic in a future lesson (on the assumption that other students might also have similar problems), using peer assistance and the use of further tests. In using peer assistance most respondents used language which implied that the assistance could come from any other student who understood it but some assumed that it would be 'brighter' or 'more able' students who would be used which implies a perception that these students do not suffer from confusion in the same way. The language used in this sort of example is telling. R18, for example, said: "Maybe – get one of the brighter pupils in the class to explain the issue. Often peer led explanations are more effective than teacher led". R13 said they would get "...stronger pupils to explain the concept".

The use of tests as a means of checking understanding was cited by R18. Testing using 'end of topic', 'end of unit' and 'end of year' assessments seemed to be a particular focus of this respondent which implies that ensuring students get grades is an important element in their teaching. The extent to which repeated testing helps alleviate confusion over a key concept could be doubted but this respondent suggests this is a potential solution to the problem.

Some respondents noted that their response to the scenario presented would be dictated by the policy established in the school in which they teach. R1 answered by assuming the scenario implied the evidence of the confusion arose in written work which is then subject to a school policy referred to as a "target forward policy". Initially, students receive written feedback and are given the opportunity to respond and if this does not resolve the issue then lunch-time revision sessions are held at which students can get one-to-one tuition. If this is also insufficient then a weekly 'drop-in' session is available. R12 also used written comments on the student's work as the first response and noted that "...students respond to my marking in line with the school policy and usually if there is confusion, this helps to clear it up in their books. I do this so that when they are revising they don't get confused." which puts the focus on the school policy and implies that the exam is the key focus of the teaching, reflecting C3.

One respondent, R14, specifically mentioned formative assessment as a means of addressing the confusion. Their answer focused on seeking to identify and structure the different levels of response required of the student. "I try to level my feedback to them based on where the confusion lies. If it is understanding then I will set level 1 targets to improve and focus purely on developing their knowledge by asking for, and checking corrections. Once confident that level1 knowledge is consistent then the focus moves on the ability to explain and apply knowledge (often the source of most confusion). The students will then move up the progress ladder to focus on how to analyse situations and form justified opinions. By clarifying where the confusion lies I can then focus on specific feedback and actions that will help the student".

6.3 Summary and Conclusion

Given the variety of questions, it is not surprising that a range of different approaches and answers were provided by respondents reflecting different realities. However, a feature that did arise was the focus on the first three conceptions of learning and in particular, C3. This might be expected given the fact that this sample were all involved in teaching a subject where the student is assessed by examination. There were some instances where broader conceptions of learning were given but these tended to be from the same respondents in each case.

This phase of the research suggests that teachers have a relatively limited view of learning as being associated with knowledge transfer and their teaching as being primarily associated with getting students through examinations. There was little evidence that deep approaches to learning, or transformational learning of the type associated with threshold concepts, was Page | 251

a key focus of the respondents in this sample. There was also limited evidence of consistency of understanding of what thinking in the subject means or the extent to which it was a focus of teaching and learning.
7. Discussion, Conclusions and Recommendations

7.1 Review of Outcomes of Phases

The research conducted has provided some answers to the questions which informed the three phases. Phase I showed that a sample of students scoring relatively high marks on the respective papers, could secure these marks even though there was some evidence of troublesome knowledge. Answers to higher tariff questions which addressed the command word in the question and followed a 'plan' meant that the student was able to access the higher levels of the mark scheme. In Phase II, the variations, both within the separate mark schemes and across the two, were sufficient to suggest that the sample group did not have a common understanding of the mark schemes used in this assessment exercise. Not only were there differences in the marks awarded but the qualitative element of the research also confirmed the different perceptions of the evidence of learning the teacher/markers were confronted with. There were a number of very different realities expressed in Phase III but the evidence did point to a conception of learning that was more similar to a surface and strategic approach to learning than a deep approach. The focus on knowledge, the use of tests, the adherence to lesson plans which met inspection requirements, and the importance of ensuring that student results were supportive of the targets which either school heads set or which are perceived to be set by the inspectorate/government were common features in the responses given. There were some references reflecting conceptions of learning which were characteristic of deep learning but these were in the minority. In addition, it seemed clear that there was no common understanding of what thinking in the subject meant.

The extent to which the findings of the phases in this research are generalizable cannot be determined. The samples used in each phase were small and the extent to which participants could be said to be representative cannot be considered with any certainty. However, the concern about the validity and reliability of assessments, which underpinned the research questions, is something that is of national interest. The qualifications regulator, Ofqual, published a report in 2014 which echoed some of the findings contained in this thesis. The report was a review of the quality of marking in exams including those at A-level and GCSE. The review found that "some teachers had a limited understanding of marking..." (Ofgual, 2014c, p4) and that the quality of mark schemes are essential to reliable marking. The report did note that in subject areas such as economics and business, where extended writing is a feature of the assessment, there will be "legitimate differences of opinion between equally gualified and skilled examiners" (Ofqual 2014c, p7) and that there would as a result be some variation in the marks awarded to students in examinations. Where these variations are 'in tolerance' they are considered to be acceptable and the suggestion seemed to be that there could never be any such thing as a perfect mark scheme which would reflect a common and agreed understanding of the evidence of learning. That view is supported by the evidence in this research but the question remains as to what is an acceptable level of tolerance in the variation such that the assessment can be regarded as valid and reliable?

There was evidence from Phase III that teachers in the sample use the system to inform their teaching and their approach to teaching. Participants referred to features of lesson plans which are deemed to be characteristic of 'good teaching' and how a focus on knowledge can help students achieve exam grades, particularly with reference to the way key terms are interpreted and used as part of the teaching and learning process. The Ofqual report (Ofqual, 2014c) in the section on 'The role of teachers, schools and colleges in marking' noted that some schools are focused on the grade boundaries at critical levels, such as the C/D boundary at GCSE, in relation to accountability measures imposed by government. The report observed that information sought from awarding bodies by teachers was 'generally' sought to "help them better prepare students for exams" (Ofqual, 2014c, p51). This echoes the focus on C1 - C3, which was the predominant conception identified in Phase III. Furthermore, the enquiries about results system, which aims to provide a means through which errors in marking can be rectified is,

according to the Ofqual report, being used tactically by schools and colleges to meet targets.

Phase II presented some evidence that there was not a common understanding of mark schemes and that evidence of learning presented in the form of student written answers could be interpreted in different ways by teacher/markers in the sample. Of qual makes it clear that in a subject area like economics and business, where extended writing is a feature of assessments, there are likely to be difference of opinion amongst markers (Ofqual, 2014c). It is made clear that the main function of a high quality exam system is that it should be both valid and reliable, and "must measure what they set out to measure and they must do so consistently" (Ofqual, 2014c, p11). It goes on to say that marking is "not an exact science" and especially in subject areas where extended writing is a key part of the assessment "a mark is a human judgement of a student's work and is only ever an approximation of his or her true score" (Ofqual, 2014c, p11). It is accepted that there will be some variation in the marking by individuals and that there are tolerances which help define the extent to which these judgements are acceptable. Differences in human judgement are meant to be minimised by the provision of a mark scheme, defining the skills being assessed and what they should look like, which examiners subscribe to and share a common understanding of, following standardisation. In Phase II, teacher/markers were not standardised in a way expected in an examination but the length of the mark scheme and the amount of evidence of learning presented to the teacher/markers did contain some guidance as to the skills being sought. In both the Bloom and the SOLO mark scheme, different interpretations arose. This could be to do with the quality of the mark scheme, which along with the different conceptions of learning and understanding of what is valid in a response to a particular question, could explain some of the variation in the participants' responses. The quality of the mark scheme is a fundamental aspect of valid and reliable marking according to Ofqual (Ofqual 2014c). Ofqual notes that the construction of mark schemes for assessments which involve some element of written

response (i.e. is not an objective test) is 'not an easy task' and that it is not possible to capture all possible responses. The existence of some element of subjective judgement will invariably mean there will be differences in interpretation but Ofqual suggests that: "...there is evidence that small improvements to the structure, presentation, content and wording of mark schemes could yield some of the biggest improvements in marking reliability" (Ofqual, 2014c p39). Ofqual does not specify what evidence supports this view and the comment also begs the question what else is there in a mark scheme other than structure, presentation, content and wording? It suggests a major overhaul of mark schemes is necessary.

How mark schemes are perceived is also an important element of assessment. In Phase II, the use of two different mark schemes, one more familiar to the teacher/markers than the other, presented the necessity of the participants to think about what they were looking at in the student answers in a different way. The imperative was to move from a relatively atomised approach to looking at an answer, where particular evidence of knowledge and skills are identified common to the way the Bloom-based mark scheme is used, to one where the answer has to be looked at holistically and the quality of understanding considered. The Ofqual report noted that perceptions of the mark scheme was the second biggest challenge to examiners after time pressure (Ofqual, 2014c). In commenting on mark schemes, respondents to Ofqual's survey noted that wording needed to be clearer and that mark schemes could be too vague, whereas others noted that mark schemes were too prescriptive. There were also calls for being able to more clearly distinguish between good and excellent knowledge although it was not made clear whether such a distinction would be made on the basis of the context of the skills students employed to demonstrate and use knowledge, or whether it was simply knowledge itself. The calls for clearer wording and improved ability to distinguish between students was not only from examiners but also teachers with the latter invariably noting that this would help in the preparation of students for examinations; such a response would have been categorised within conceptions 1 - 3 in Phase III of this research.

The quality of mark schemes is undoubtedly important in improving the reliability of assessment, particularly in a subject area like economics and business where extended written responses are a feature of the assessment, and where marker judgement is required. Typically exam papers have been written in a sequence. Paper setters find some stimulus material and decide on the content areas of the specification to be covered and write a set of questions covering the range of skills determined by the assessment objectives. The writing of the mark scheme then follows. Pollitt and Ahmed (2008), argue that such an approach can lead to reduced validity and reliability. Pollitt and Ahmed (2008) argue that the focus ought to be on considering the behaviours required of students. These behaviours might relate to specific content and particular skills but thinking first about, and understanding these behaviours, would put paper setters in a better position to "...be in control of the mental behaviour that occurs when a student meets a question" (Pollitt and Ahmed, 2008, p5). Pollitt and Ahmed (2008) propose a 'system' for paper setters which they call Outcome Space Control and Assessment (OSCA). Outcome spaces are the evidence of learning paper setters are seeking. In setting a question an assessor is expecting a particular response from the student. For students taking the assessment one of the skills that is actually assessed is to get inside the mind of the guestion setter to have some understanding of their thinking and expectation; this also applies to markers in looking at the mark scheme and trying to interpret it.

OSCA is based on the assumption that any question will elicit different types of responses. These responses might be ones anticipated and expected by the question setter and range from good to poor. However, a question will elicit responses from students that the question setter did not expect or anticipate and there will be occasions where the expected answers that the question setter anticipated will not be in evidence. In these cases the mark scheme may not adequately cover the range of responses and require the question setter to have to amend the mark scheme. This invariably happens following standardisation meetings at GCSE and GCE after the senior examining team have had a chance to look at candidate responses. Examiners, therefore, get Page | 257 the original draft mark scheme which they are expected to look through and begin to internalise but some time later (usually a week) have to unlearn the original mark scheme and instead internalise a new mark scheme, which could look very different to the original. If there is evidence of outcomes that are not desired or expected then question setters have "...lost control of the students' thinking processes" (Pollitt and Ahmed, 2008, p7).

To overcome this, Pollitt and Ahmed suggest that turning the paper setting process on its head might lead to an improvement in the quality of mark schemes, better questions and as a result, more reliable assessments. Pollitt and Ahmed (2008) suggest that the question writing process begins with clear thinking about what evidence of learning is required and what 'good' and 'poor' evidence of learning will look like. This allows the paper setter to be specific about the evidence they wish to elicit enabling them to discriminate between students - the 'desired outcome space'. When the assessment setter has thought this through, the next stage in the process is to articulate the desired outcome space in the form of the mark scheme. At this point, the wording of the questions which will best elicit the evidence of learning desired, can be formulated. Pollitt and Ahmed believe this process would improve the chances of a valid assessment through: "...eliciting evidence of the right kinds of mental behaviour - the things we want them to show us they can do - and by evaluating the resulting performances in order to make valid inferences about competence." (Pollitt and Ahmed, 2008, p8).

7.2 Aims and Objectives and Assessment Objectives

The focus on mark schemes is an important element in improving reliability and validity in assessment. The comments noted in the report by Ofqual and approaches such as those by Pollitt and Ahmed, can be taken into consideration but ultimately these comments and observations are being made within the context of the status quo. Assessment at GCSE is defined by the Subject Criteria which is published by Ofqual. The Subject Criteria for business subjects lays out the aims and learning outcomes, assessment objectives, scheme of assessment and grade descriptions for the subject area.

This provides the base upon which the awarding organisations develop their specifications. The existence of the subject criteria means that there is some degree of consistency between the offerings of different awarding bodies, whilst also providing opportunity for differentiation.

The GCSE subject criteria is also open to a degree of interpretation by the awarding bodies in the way in which they create the detail of their specifications. In 2011, the subject criteria for GCE advanced level in some subject areas, including economics and business, were revised in preparation for the new specifications for first teaching in September 2015. The aims and objectives in the subject criteria for economics made specific reference to thinking in the subject: "AS and A level specifications in Economics should encourage students to:...use an enquiring, critical and thoughtful approach to the study of economics and an ability to think as an economist..." (Ofqual, 2011a, p3). Interestingly, the requirement to think in the subject was not included in the aims and objectives of the subject criteria for business subjects at A level.

The Subject Criteria gives no definition to what 'thinking as an economist means' but this requirement, along with the other aims and objectives, underpin the assessment. In the discussion of the purpose and nature of assessment and assessment objectives in Chapter 2, assessment was defined as the gathering of evidence of the desired measure of learning and the subsequent interpretation about that evidence and its use, to make judgements about students' achievements following a course of instruction (Harlen, 2007). Learning was defined as a change in thinking as a result of the acquisition of new knowledge, skills and understanding. The link to threshold concepts in this thesis is where changed behaviour will mean students have a new and altered view of the world around them that is transformative - they have progressed through the threshold and now think in a very different way about the subject matter. What defines this different way of thinking is the way in which economists and business people think and so thinking in the subject has to be at the core of the aims and objectives of the learning. This

thesis is making a statement that the purpose of the assessment, what we want students to learn and evidence, is that they are beginning to think in the subject. The question then arises, what is the best way of achieving this aim?

This requires a consideration of what thinking in the subject actually means and getting some common agreement amongst teachers so that the direction of learning and the way in which teaching and learning programmes are based, develop and encourage this in students. The assessment measures the extent to which students show their ability to think in the subject. Further clarification of thinking in the subject will be given in the next section.

The subject criteria as it currently exists, provides two, not necessarily complementary, sets of aims: the stated 'Aims and Learning Outcomes' and the assessment objectives. The extent to which the aims and learning outcomes stated in the subject criteria can actually be measured is open to some debate.

The aims and learning outcomes are as follows:

"GCSE specifications in business subjects must enable learners to:

- actively engage in the study of business and economics to develop as effective and independent learners, and as critical and reflective thinkers with enquiring minds;
- use an enquiring, critical approach to distinguish facts and opinions, to build arguments and make informed judgements;
- develop and apply their knowledge, understanding and skills to contemporary issues in a range of local, national and global contexts;
- appreciate the range of perspectives of different stakeholders in relation to business and economic activities;
- consider the extent to which business and economic activity can be ethical and sustainable." (Ofqual, 2011b, p3)

Given the definition of assessment used in this thesis, it is difficult to see how some aspects of these aims and learning outcomes could be measured. For

example, can an assessment accurately measure the extent to which 'learners' are enabled to 'actively engage in the study of business and economics'? Can the answer to questions, similar to those used in the first two phases of this research, measure the extent to which a student is *actively* engaged? Can such guestions show the extent to which a student has developed as an 'effective and independent learner' (whatever an 'effective learner' means)? Do the answers to the questions allow an examiner to make inferences about the extent to which a student is a critical thinker and has an enquiring mind? An assessment may include questions which cover certain local, national and global issues but does this the mean that the student is able to apply that knowledge and understanding to other contexts? Do the sort of questions highlighted in the first two phases of this research enable the measurement of whether a student is able to use an 'enquiring and critical approach to distinguishing between facts and opinion' and can continue to do so when presented with other contexts outside the confines of an examination paper? Does the assessment enable inferences to be made about the student's ability to make informed judgements across a range of different contexts associated with economics and business?

It can be argued that questions can be devised which do measure whether a student is able to recognise different perspectives of stakeholders and present those in an answer. Students might also be able to note instances where business and economic activity are focused on ethical and sustainable principles but can an assessment consider the extent to which business and economic activity can be ethical and sustainable?

The subject criteria then provides a further focus in the form of the assessment objectives. For awarding organisations and examiners, it is the assessment objectives that drive the purpose and structure of the assessment. The proportion of weighting of the exam papers in GCSE economics and business, and indeed all other GCSEs and GCEs, are based on the assessment objectives and not on the aims and objectives. Nowhere, in any specification, is there any reference to the weighting of an exam paper in respect of how

far it 'engages students in the study of business and economics' nor any of the remaining aims and learning outcomes. It is likely that in focusing on the higher order skills of analysis and evaluation that some students may develop a more critical approach to a question and might be in a position to build an argument and make informed judgements but how the two are related and reflected in the assessment is not made clear.

Phase III of the research highlighted the focus which participants in this sample had on the exam, on grades, on knowledge and on meeting the targets set by schools external bodies such as the government and Ofsted. It is not surprising that teachers may focus on the assessment objectives which identify and make clear how marks can be awarded for answers and help students get certain grades. The way in which the subject criteria places the emphasis on assessment objectives means that wider benefits of the assessment in encouraging thinking in the subject and deep learning are sacrificed.

The focus on assessment objectives has implications for the mark schemes, which must reflect the proportions of knowledge, application, analysis and assessment, and the extent to which these are measured in different questions. In the methodology chapter, the way in which a question can be broken down into the different assessment objectives and be the subject of disagreement was highlighted. The interpretations of the question setter in that mini-pilot of the breakdown of assessment objectives in a question compared to that given in the original paper, further emphasises how the accuracy and reliability of any such breakdown can be subject to question. If the allocation of assessment objectives to a question is unreliable then the mark scheme, which is acknowledged as being a central element in the reliability of marking, is also questionable. When this is added to the different interpretations of mark schemes by teacher/markers, which Phase II of this research indicated was a factor, the ability of the assessment to deliver a measure of the evidence of learning which meets the aims and objectives and/or the assessment objectives in a clear and unequivocal

fashion is at issue.

7.3 Thinking in the Subject

If the aims and objectives provided by the subject criteria give conflicting and uncertain measures of learning an alternative focus of the desired learning outcome may be needed. An option is to have thinking in the subject as the primary learning outcome of the assessment. Thinking in the subject reflects the definition of learning as a change in thinking and incorporates deep learning and the role of threshold concepts and troublesome knowledge. If thinking in the subject was the primary desired learning outcome, the assessment can be designed to measure the extent to which the evidence of learning matches this aim.

The use of OSCA (Pollitt and Ahmed, 2008) would be relevant as a way in which the assessment design process could be informed. The focus has to be on the quality of learning outcome and not on separate elements of knowledge and skills which tend to be the consequence of using an assessment based on Bloom's taxonomy. Pollitt and Ahmed (2008, p3) comment that: "an exam can only contribute to valid assessment if the students' minds are doing a representative sample of the things we want them to show us they can do; and if we give credit for, and only for, evidence that shows us they can do it.". Any assessment will not be able to cover all the 'things we want students to do' (i.e. think in the subject) and so what students show they can do must be representative of the overall aim. The mark scheme must be designed to ensure that what is being credited is linked to the primary aim of the assessment and when students provide evidence of learning of this, that markers recognise it and reward it appropriately.

This rests on a major assumption that the primary aim can be agreed upon. In Chapter 2, an outline of thinking in the subject was provided. Thinking in the subject reflects a change in thinking linked to the definition of learning given and to deep learning. Deep learning (an intention to understand, looking for connections and meaning, internalising meaning (changing behaviour) and constructing new realities) is reflected in the SOLO taxonomy at Level 5. Evidence of learning would be classed as extended abstract where students demonstrating this level of learning would have the ability to make connections both within the bounds of the subject and outside it, and be able to take their understanding and make generalisations to other contexts. The link between deep learning, thinking in the subject and threshold concepts is clear in reflecting the change in thinking which is the focus of the definition of learning used in this thesis.

If thinking in the subject was to be the primary learning outcome there has to be a common understanding of what thinking in the subject means and a shared conception of learning by teachers. In Chapter 2 it was noted that a number of undergraduate textbooks in economics have an introductory chapter which details what thinking as an economist means. Typically, these introductions include understanding of core concepts such as choice, tradeoffs, opportunity cost and the margin and an awareness of the role of incentives, how models are used to predict behaviour and the role of assumptions in using models. These could be argued to be at the heart of thinking as an economist and would be recognised by anyone who is an 'expert' or is deeply associated with the subject area.

As has been highlighted in this thesis, human beings have different realities and despite literature on the methodology of economics (Blaug, 1992) it cannot be assumed that there is this common understanding of what thinking in the subject means. I was able to experience this in a conversation with a colleague during meetings of an accreditation panel for new GCE A-level qualifications in England and Wales in the summer of 2014. As noted above, the new subject criteria for economics A-level includes reference in the aims and objectives to 'thinking as an economist'. The panel was debating the extent to which the submissions being considered met this criteria and it became clear that thinking in the subject was not something that could be assumed to be commonly understood or accepted.

Subsequent discussions with one of the panel members (Cafferty, 2014) highlighted some central points for debate. One of these points was the

context under which one might be expected to think as an economist. Economists exist in many different walks of life, there are professional economists who are employed by banks, financial institutions, labour and welfare organisations and public bodies, there are academic economists who immerse themselves in research, those who focus on teaching and economists who work for government departments or who are advisors or consultants to large public limited companies. There are also students taking examinations in the subject area at different levels from GCSE to post-graduate levels. Each of these represents a different context under which individuals will be expected to think as economists. In each context, however, are the core concepts and principles, as outlined in the first chapter of introductory texts, at the heart of the way in which economists in these different contexts 'think'?

Cafferty (2014), suggests that the core concepts and principles have to be allied to the requirement to use a particular methodology. The methodology of economics is based partly on scientific method, of developing theories and models which are used to predict and explain and are underpinned by systematic gathering of evidence and classification of knowledge (Blaug, 1992). Unlike the natural sciences, economics is a social science in that it deals with human beings. There are theories of human behaviour which have pervaded economics for many years. These are "...rigorous, deductive theories of human action that are almost wholly lacking in...other behavioural sciences" (Blaug, 1992, pxxv). There has been much criticism of these theories of human behaviour and in particular research by psychologists like Amos Tversky and Daniel Kahneman (2008), has shed light on what has come to be called behavioural economics. Much of this research casts doubt on the assumptions of rational behaviour that underlie key models in economics.

There are also other branches of economics which need to be considered. A significant element of the current specifications in economics and business, at both GCSE and GCE, are based on neo-classical paradigms and these have been challenged by behavioural economists, Marxist economics and the

Austrian school. Cafferty (2014) notes does this mean that Marxist and behavioural economists do not think like economists?

In thinking as an economist or business person the response to choices and decisions made is central. Decision making is fundamental to both these subject areas and regardless of the paradigm, whether it be neo-classical, Marxist or Austrian, human beings are faced with the fundamental problem of scarcity of resources and the need to make choices of how best to make use of these scarce resources in satisfying wants and needs. The focus of the teaching in economics and business is about making decisions and the thinking which goes into making a decision. This thinking will include concepts such as opportunity cost, the margin and trade-offs and will use models to help structure thinking and predict the consequences of making decisions. Thinking in the subject also requires students to be aware of the assumptions that are made in constructing models and the limitations of theories that have been developed and of competing explanations of behaviour. This is as much a part of thinking in the subject as being able to manipulate a supply and demand diagram. Unlike the natural sciences, economics and business rarely has any 'right' answers and it is imperative that teachers help students to understand this and that any decision will have costs and benefits.

The two questions used in Phase II of the research can highlight these points. In Question 1, the student was asked to '*Explain two reasons why a business such as Nokia might use the Boston Matrix*'. At the heart of this question is the understanding that Nokia has to make decisions about its product lines and portfolio and the Boston Matrix (whether one agrees with its assumptions and premise or not) is a way in which it can quantify and qualify this decision making. Students need to be aware that the business is making decisions based on an evaluation of costs and benefits and that opportunity cost, the margin, assumptions about how humans behave and trade-offs will all have to be factored into its decision making. Equally there will be external factors which Nokia will have to consider such as the political, social and ethical environment and it may be that whilst these are not directly asked for in this particular question, they would be part of the curriculum and legitimate candidates for other types of question, which may or may not, involve an awareness of the Boston Matrix.

Question 2 was 'Using all the evidence and your knowledge of business and economics, assess the strength of the case for hosting the Olympic Games in London in 2012'.

Decision making is the basis of this question also. An understanding of opportunity cost would be a pre-requisite of the basis of any answer but in addition, candidates might also be expected to take into consideration the fact that economists and business people need to consider wider factors in arriving at a decision of whether to host the Games. The real cost of hosting the Olympic Games (the private and social costs) might be higher than the immediate economic benefits in the form of ticket, media and sponsorship revenue and so on but decision makers might need to accept that there are wider social benefits that need to be considered such as the longer term economic benefits of the regeneration of the area of East London in which the Games were held, the legacy to the community, the external benefits of improvements in infrastructure both to the region itself and to the wider economy in the South East. There might also be an intangible 'feel good' factor which decision makers would take into consideration.

All these can be seen as part of the thinking which would be expected of someone versed in economics and business as opposed to those not familiar with the subject. Ultimately, teachers are encouraging students to understand that decision-making requires the consideration of costs and benefits, both qualitative and quantitative. Some financial and commercial considerations might relatively easily be quantified but there are a number of qualitative factors not so easy to quantify that will impact on decision making and may be deemed more important than the financial considerations. This approach and this way of thinking is characteristic of economics and business and is employed in many different contexts. One of the main concerns raised by Cafferty (2014) was knowing when people think as economists. It can be argued that economists apply their training and way of thinking in many different contexts and it is not possible to specify exactly when this context is appropriate and when it is not.

This thesis is about GCSE economics and business and so the context is the application of this way of thinking to the content which is deemed to be appropriate and part of at this level. The basis of the curriculum content is determined by the subject criteria referred to above. It is debatable whether the content outlined in the subject criteria is appropriate; is the expected content too extensive? Does it allow teachers to foster the sort of understanding and thinking suggested in this thesis? Are quality and understanding sacrificed for quantity? These are questions which decision makers in the Department of Education and Ofqual make in consultation with teachers and interested parties. Given the subject criteria and the resulting specifications produced by the awarding bodies, teachers are faced with interpreting and seeking to impart the knowledge and skills outlined in the specification to students. If the aims and learning objectives of the specifications are confused then the resulting teaching will also be confused. If the main aim is to encourage students to think in the subject, and the assessment is focused on measuring the extent to which students demonstrate evidence of learning that they can do this, then the way of teaching and the conceptions of what learning is about in this particular context will follow accordingly. In summary, if the main aim of the specification is to encourage students to think in the subject then it is expected that students will use the tools, models, concepts and methods used in the subject area to consider the relative costs and benefits of decision making, to understand how and why decisions are made and make judgements on decision making which reflect an understanding of these tools, concepts and models, their assumptions and their limitations.

7.4 Mark Schemes and Assessment

If thinking in the subject is the basis of the specification then the assessment has to be designed to measure the extent to which students demonstrate that they can use the tools, models, concepts and methods to explain decision making and to make judgements which reflect the use of the tools etc. This thesis has presented evidence that the current assessment regime based on Bloom's Taxonomy raises questions about whether students really understand the content they are dealing with. Despite getting high marks in exam papers there are instances where troublesome knowledge appears prevalent. Teacher/markers in the sample in this research demonstrated varied interpretations of mark schemes and viewed evidence of learning in relation to different mark schemes in different ways. Teachers in Phase III showed some evidence of having conceptions of learning that were redolent of surface and strategic learning rather than deep learning. The assessment regime plays some part in the formation of these conceptions.

To change conceptions and the approach to teaching and learning, a focus on encouraging thinking in the subject is a starting point but this must be supported by a change to the assessment regime. The approach to setting questions suggested by Pollitt and Ahmed (2008) may be one way to begin to achieve such a change but Pollitt and Ahmed's system is still based on an assumption that the overall regime reflects Bloom's Taxonomy. Pollitt and Ahmed (2008) note that Bloom's Taxonomy is 'unsuitable' as it does not reflect the cognitive processes which might be deemed desirable as the basis of the learning students' exhibit. Pollitt and Ahmed (2008) make reference to Peel's single hierarchy (Peel, 1971), which classifies thinking characterised by 'mentioning, describing, explaining and extended explaining'. They also note that this approach was developed by Biggs and Collis (1982) in the form of the SOLO Taxonomy.

The aims and learning outcomes specified in subject criteria must be the basis on which the assessment is designed. Separating out the knowledge and skills required from the more general aims and learning outcomes gives rise to the confusion over the focus of the assessment which this thesis is arguing exists at present. Instead, the aims and learning outcomes would be the assessment objectives. These would provide an outline of the knowledge and skills which help to encourage thinking in the subject and the measure of evidence of learning to this end. If the purpose of the assessment is to measure the extent to which students demonstrate evidence that they can 'do what we want them to do' then it has to be made clear in the aims and objectives what it is that is required of students.

The aims and learning outcomes might look something like the following: Assessment of GCSE Economics and Business will result in learners being able to:

- Show an ability to think in the subject by employing the tools, models, concepts and methods used by economists and business people to address problems and issues and make decisions in the context of scarce resources and unlimited wants and needs;
- Demonstrate an understanding of the role and importance of evaluating costs and benefits in making decisions;
- Show an understanding of a range of specified concepts and be able to apply these concepts to explain economic and business issues and show how economists and business people approach decision making and problem solving;
- Demonstrate an appreciation and awareness of the assumptions underlying the tools, models, concepts and methods used in economics and business and the limitations which arise as a result;
- To critically examine and explain economic and business methods and the approaches to dealing with problems, issues and decision making and be able to critically analyse and assess proposed solutions;
- Think in terms of alternatives and analyse and evaluate the quantitative and qualitative factors which inform decision making.

If these are the assessment objectives - what the desired learning outcomes following a course of study in the subject are - the focus of the assessment is on how to measure the extent to which students demonstrate these outcomes. The subject criteria can include a list of content which it is deemed desirable to cover over the course of study. This content might include the 'traditional topics' which make up most economics and business specifications. Given the assessment objectives, it is desirable that quality is preferred to quantity and that the range of content is more limited than that presently making up awarding body specifications. The focus becomes encouraging students to explore a wide variety of different contexts within a topic area to build up an understanding of the way in which thinking in the subject is relevant to these contexts. For example, if the working of markets is a central topic in the content, teachers need to have the time to use realworld examples in a variety of different contexts to help students understand the model of supply, demand and the price mechanism, the assumptions which underpin the model, to investigate and explore instances where the model works as predicted but also the instances where the model does not work and where the assumptions and limitations in the model mean decision making and outcomes might be different to that anticipated or predicted.

The mark schemes for the assessment are developed to specify the knowledge and skills that are desired which provide evidence of the learning in relation to the aims and objectives specified. Equally, the mark scheme must be clear on how marker can view the evidence of learning to differentiate between students. The way the mark scheme is written incorporates and describes the way in which markers can recognise what knowledge is appropriate and expected, what the appropriate and expected application of this knowledge will look like and what analysis and evaluation skills will look like. For example, at a very low level, students who make judgements but without any further support will gain some reward but those who provide one or more supporting statements will gain more marks and those that are able to provide balanced supporting statements, reflecting different sides to the argument, would be providing evidence at the top level. In demonstrating analysis skills, markers might be directed to look for examples where students make connected statements which represent the construction of an argument which provides evidence of critical examination or which shows how a topic is Page | 271 being broken down into its constituent parts. The use of words or phrases such as, 'because...', 'this leads to...', 'this results in...', 'as a result...', would be indicators of where connected statements are being made. How frequently they are made would be an indicator to help the marker determine the appropriate level in which to place the answer. The level would be defined by the number of connected statements with higher levels reflecting more connected statements and the sophistication of the connections made.

If mark schemes are devised in this way then the way in which the assessment is marked will move away from an award of a numerical mark, totalled and subjected to the sort of parametric statistical tests referred to in Phase II of this study. Phase II indicated that a reliance on the assumption of a common understanding of mark schemes and marks awarded, which would be necessary to enable parametric statistics to be used with confidence, is open to question. Placing student answers in levels and relating these to an overall levels scale linked to the assessment objectives along the lines of that used in the SOLO Taxonomy would give an indication of the quality of the learning outcome and be more closely aligned to the assessment objectives as now stated. This would replace the current system whereby the grade descriptors offer yet another measure of student performance in addition to the assessment objectives and aims and objectives. One major challenge of this approach would be in ensuring that teachers became familiar with the levels and what evidence of learning looks like in each level. This challenge can be met by careful wording of mark schemes, clear descriptions of what is expected in each level in the mark schemes and a sufficient number and variety of examples of evidence of learning to enable teachers to develop familiarity and confidence in interpreting the levels.

7.5 Limitations and Recommendations for Further Research

The main limitation in this thesis is the small sample sizes used in each of the three phases. Whilst a small sample size does not in itself mean the research is any less trustworthy, the ability to generalize and argue that the findings are representative of student performance in other economics and business

examinations or across other subjects, of other teacher/markers or of teachers' conceptions of learning is a limitation. In Chapter 3 it was noted that qualitative research of the type, which is used to different degrees in each phase of this research, has to have the characteristic of being trustworthy. Each phase was subject to a rigorous approach to the method employed and within the framework of a constructivist paradigm the research represents a reality and some conclusions which reflect one reality.

It has to be accepted that there are other realities and in particular the focus of this research on economics and business may mean that the applicability of any of the conclusions are not relevant, or would only be so, if applied to research in each subject area.

This research used a particular GCSE qualification in economics and business which stands apart from typical economics and business qualifications at this level. The qualification developed in the early 1990s from a belief that the two disciplines were interrelated and that combining the two areas of study would provide for the increasing economic and business literacy needs of students who end up working in industry and commerce. At the time there was only one qualification which combined the two subject areas, partly because study of economics and business at this level is optional and the market relatively small, and partly because many teachers do not feel that they have the expertise to cover two different subject areas. Teachers trained in economics might feel uncomfortable having to teach more business focused content such as marketing whereas business studies trained teachers might feel equally uncomfortable teaching some of the more analytical aspects of economics - the market mechanism and market structures, for example. In the next few years it is likely that GCSE economics and business will be subject to a similar review and update which has been witnessed at GCE A-level. Whether there is a GCSE in economics and business in the future is uncertain and it may be that some awarding bodies opt to offer either a GCSE in business studies or a GCSE in economics but not both. This research is relevant to either economics or business studies and the idea of thinking in

the subject applies to both. The way in which the subject criteria and assessment objectives are generated could be adapted to each subject area if future GCSE reform led to qualifications being offered in the two subject areas but not as a combined qualification.

Further research exploring the extent to which a common understanding of what it means to think in a subject is necessary. Getting a common understanding is essential if it were to form the basis of the aims and learning outcomes of any future subject criteria. In Phase III, the agreement amongst participants of what key economic and business concepts should form the basis of a teaching and learning programme in the subject area was limited. The range of different concepts cited was extensive and the lack of any notable frequency of agreement might be peculiar to this sample of the population of economics and business teachers. If this sample were to be taken as being representative, however, it would suggest that getting agreement on what constitutes thinking in the subject means and what it looks like in evidence of learning would be challenging.

Ofqual is undertaking a review of mark schemes (Ofqual, 2014c) and it is the case that there is much research which could be done in this area. In particular, the generation of mark schemes using OSCA and the way in which specific skills are identified and rewarded to ensure greater clarity, would lead to marking becoming more reliable. The focus on mark schemes in the Ofqual report (Ofqual, 2014c) does suggest that this is an area of concern and if assessments at national level are to gain more credibility and teachers and wider stakeholders are to have more confidence in the assessment system, research into ways of improving mark schemes is essential.

One of the underlying themes of this thesis has been that assessment can drive the teaching and learning process. Phase III highlighted possible links between the assessment regime and conceptions of learning which were more focused on surface and strategic learning than deep learning. If teachers are to be encouraged to approach their teaching and learning in a way which reflects deep learning more effectively then the assessment has to provide the basis of that change. Further research on whether a change in emphasis of the assessment objectives, as proposed in this thesis, would lead to a different approach by teachers and whether students would emerge from two-year courses equipped with skills helping them think in the subject and adopt more of a deep learning approach is necessary. In the absence of a national change to the assessment objectives in the way proposed, it would be difficult to see how such a change could be researched. If a pilot was run in a sample of schools there would be a number of ethical and practical hurdles to overcome.

8. Bibliography

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruickshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., Wittrock, M.C. 2001: *A Taxonomy for Learning, Teaching and Assessing: A revision of Bloom's taxonomy of educational objectives.* London: Addison Wesley Longman.

Aristotle. 2009: *The Nicomachean Ethics (Oxford World's Classics)*. Translated by D. Ross and revised by L. Brown. Oxford: Oxford University Press.

Arnold, R.A. 2005: *How to Think Like an Economist.* Mason, Ohio: South Western, Cengage Learning.

Ashwin, A.K. and Pitts, K. 2007: Developing learning materials for the unknown online learner. In M.K. McCuddy., H van den Bosch., W.B. Martz Jr., A. V. Matveev. and K.O. Morse (eds.), *The Challenges of Educating People to Lead in a Challenging World.* Dordrecht, Netherlands: Springer. **10**: pp. 273-290.

Ashwin, A.K. 2008: What do students' examination answers reveal about threshold concept acquisition in the 14 – 19 age groups? In R. Land., J.H. Meyer. and J. Smith (eds.) *Threshold Concepts Within the Disciplines*. Rotterdam, Netherlands: Sense Publishers.

Atherton, J.S. 2013: *Learning and Teaching: SOLO taxonomy*. 10th February 2013. Accessed 12/3/14.

http://www.learningandteaching.info/learning/solo.htm

Ausubel, D.P., Novak, J.D. and Hanesian, H. 1978: *Educational Psychology: A cognitive view*. New York: Holt, Reinhart and Winston.

Baillie, C., Bowden, J.A. and Meyer, J.H.F. 2013: Threshold capabilities: Threshold concepts and knowledge capability linked through variation theory. *Higher Education*, **65**: pp. 227-246.

Bandura, A. 1977: *Social Learning Theory*. Englewood Cliffs, New Jersey: Prentice Hall.

Barradell, S. 2013: The identification of threshold concepts: A review of Page | 276

theoretical complexities and methodological challenges. *Higher Education*, **65** (2): pp. 265-276.

Beinhocker, E. D. 2013: Reflexivity, complexity and the nature of social science. *Journal of Economic Methodology*, **20** (4): pp. 330-342.

Bell, R.M. 1979: Building an experiential approach to learning. In N.W. Steinaker, and M.R. Bell. (eds). *The Experiential Taxonomy: A new approach to teaching and learning.* New York: Academic Press.

Bigge, M.L. and Shermis, S.S. 1999: *Learning Theories for Teachers 6e.* Reading, Berkshire: Addison Wesley Longman.

Biggs, J. and Collis, K. 1982: *Evaluating the Quality of Learning: The SOLOtaxonomy*. New York: Academic Press.

Biggs, J.B. 1987: *Student Approaches to Learning and Studying*. Hawthorn, Victoria: Australian Council for Educational Research.

Biggs, J.B. 1996: Enhancing teaching through constructive alignment. *Higher Education*, **32**: pp. 1-18.

Biggs, J. 2003: *Teaching for quality learning at university: What the student does 2e.* Buckingham: The Society for Research into Higher Education and Open University Press.

Biggs, J. and Tang, C. 2007: *Teaching for Quality Learning at University 3e.* Maidenhead, Berkshire: Society for Research into Higher Education and Open University Press.

Black, P. 1999: Assessment, learning theories and testing systems. In P. Murphy (ed.) *Learners, Learning and Assessment*. London, Sage, pp. 118-132.

Black, P. and Wiliam, D. 1998: Assessment and classroom learning. Assessment in Education: Principles, policy and practice, 5 (1): pp. 7-73.

Black, P. and Wiliam, D. 2003: 'In praise of educational research': Formative assessment. *British Education Research Journal*, **29** (5): pp. 623-637.

Black, P. and Wiliam, D. 2006: The Reliability of Assessments. In J. Gardner (ed.), Assessment and Learning. London: Sage, pp. 243-263.

Blaug, M. 1992: *The Methodology of Economics: Or how economists explain.* Cambridge: Cambridge University Press.

Bloom B. S., Engelhart, M.D., Furst, E. J., Hill, W. J. and Krathwohl, D.R. 1956: *Taxonomy of Educational Objectives. The classification of educational goals. Handbook I: The Cognitive Domain.* London: Longman.

Bogdan, R.C. and Biklen, S.K. 1998: *Qualitative Research in Education: An introduction to theory and methods 3e*. Needham Heights, Massachusetts: Allyn and Bacon.

Bouchard, T. J. Jr. 1976: Field research methods: Interviewing, questionnaires, participant observation, systematic observation, unobtrusive measures. In M.D. Creswell, J.W. 2011: *Designing and Constructing Mixed Methods Research 2e.* Thousand Oaks, California: Sage.

Boulton-Lewis, G.M. 1995: The Solo taxonomy as a means of shaping and assessing learning in higher education. *Higher Education Research and Development*, **14** (2): pp. 143-154.

Boulton-Lewis, G.M. 1998: Applying the SOLO taxonomy to learning in higher education In: B, Dart and G.M. Boulton-Lewis (eds). *Teaching and Learning in Higher Education*. Melbourne: ACER, pp. 201-221

Bowden, J.A., Dall'Alba, G., Martin, E., Laurillard, D., Marton, F. and Masters, G. 1992: Displacement, velocity and frames of reference: Phenomenographic studies of students' understanding and some implications for teaching and assessment. *American Journal of Physics.* **60** (3): pp. 262 – 269.

Bowring-Carr, C., and West-Burnham, J. 1997: *Effective Learning in Schools*. London: Financial Times.

Boyett, J., and Boyett, J., 1998: *The Guru Guide: The best ideas of the top management thinkers*. New York: John Wiley.

British Educational Research Association (BERA) 2004: *Revised Ethical Guidelines for Educational Research, 2004*. Accessed 3/12/14

https://www.bera.ac.uk/wp-content/uploads/2014/02/ethica1.pdf

Bruner, J. S. 1966: *Toward a Theory of Instruction*, Cambridge, Massachusetts: Belkapp Press.

Bruner, J. S. 2006: In Search of Pedagogy: The selected works of Jerome Bruner 1979 - 2006. Abingdon, Oxon: Routledge.

Burnett, P.C., Boulton-Lewis, G.M. and Campbell, L. 1996: *Investigation of Students and Teachers' Conceptions of Learning and their Effect on the Teaching and Learning Process: A report to the schools*. Technical Report, The Research Concentration in Cognition, Learning and Development, Queensland University of Technology: Australia.

Cacioppo, J. T. and Freberg, L.A. 2013: *Discovering Psychology: The science of the mind.* Belmont, California: Cengage Learning.

Cafferty, P. 2014: Personal correspondence, July, 2014.

Carey, S. 1991: Knowledge acquisition: Enrichment or conceptual change? In S. Carey and R. Gelman, (eds.). *The Epigenesis of Mind: Essays on biology and cognition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, pp. 257-291.

Chan, C.C., Tsui, M.S., and Chan, M.Y.C. 2002: Applying the Structure of the Observed Learning Outcomes (SOLO) taxonomy on student's learning outcomes: An empirical study. *Assessment and Evaluation in Higher Education*, **27** (6): pp. 511-527.

Cohen, L., Manion, L., and Morrison, K. 2001: *Research Methods in Education 5e.* London: Routledge Falmer.

Collins, A.M. and Loftus, E.F. 1975: A spreading-activation theory of semantic processing. *Psychological Review*. **82**: pp. 297–312.

Couper, M.P., Kapteyn, A., Schonlau, M. and Winter, J. 2007: Noncoverage and nonresponse in an internet survey, *Social Science Research*, **36**: pp. 131-148.

Cresswell, J.W. and Miller, D.L. 2000: Determining validity in qualitative inquiry. *Theory Into Practice*, **39** (3): pp. 124-131.

Cresswell, J.W. 2012: *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Thousand Oaks, California: Sage.

Cresswell, J.W. 2015: A Concise Introduction to Mixed Methods Research. Thousand Oaks, California: Sage.

Curry, L. 1983: *An Organisation of Learning Styles Theory and Constructs*. ERIC document no. ED 235185.

Darling-Hammond, L., Austin, K., Orcutt, S and Rosso, J. 2001: *How People Learn: Introduction to learning theories.* (27/12/01). Stanford University School of Education. Accessed 24/2/14

www.stanford.edu/class/ed269/hplintrochapter.pdf

Dart, B., Pillay, H. and Burnett, P.C. 2000: Australian and Filipino Students' Approach to Learning; conceptions of learning and learner self-concepts: A cross-cultural comparison. *Educational Research Journal*, **15**: pp. 143-166

Davies, P. 2003: *Threshold Concepts: How can we recognise them*? Paper presented at the European Association in Learning and Instruction (EARLI) Conference, Padova, August 26 – 30th 2003. Accessed 18/12/13

http://www.staffs.ac.uk/schools/business/iepr/docs/etcworkingpaper(1).doc

Davies, P. and Brant, J. 2006: *Business, Economics and Enterprise: Teaching school subjects 11 – 19.* Oxford: Routledge.

Davies. P. and Mangan, J. 2005: *Recognising Threshold Concepts: An exploration of different approaches.* Paper presented at the European Association in Learning and Instruction (EARLI) Conference, Nicosia, August 23 - 27th 2005. Accessed 16/01/06

http://www.staffs.ac.uk/schools/business/iepr/info/Economics(2).html

Davies, P. and Mangan, J. 2006a: *Trajectories of Students' Learning: Threshold concepts and subject learning careers.* Paper presented at the annual conference of the Society for Research in Higher Education, University of Brighton, 13th - 15th December 2006. Accessed 23/10/13

http://www.staffs.ac.uk/schools/business/iepr/etc/WorkingPapers/etcworki

ngpaper4.pdf

Davies, P. and Mangan, J. 2006b: *Embedding Threshold Concepts: From theory to pedagogical principles to learning activities*. Paper presented at the Threshold Concepts Symposium, Glasgow, September 1st 2006. Accessed 14/6/13

http://www.staffs.ac.uk/schools/business/iepr/etc/WorkingPapers/etcworkingpaper3.pdf

Denscombe, M. 2008: Communities of practice: A research paradigm for the mixed methods approach. *Journal of Mixed Methods Research.* **2** (3): pp. 270 – 283.

Denzin, N.K. and Lincoln, Y.S. 1998: *The Landscape of Qualitative Research: Theories and issues.* Thousand Oaks California: Sage.

Donche, V., De Maeyer, S. and Van Petegem, P. 2007: *Teachers' Conceptions* of Learning and Teaching and their Effect on Student Learning. Paper presented at the British Educational Research Association (BERA) Conference, Institute of Education, University of London, 5 - 8 September 2007. Accessed 13/8/13. www.leeds.ac.uk/educol/documents/166752.doc

Dorman, J.P. and Knightly, W.M. 2006: Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks. *Educational Studies*, **32** (1): pp. 47-58.

Dunnette, M.D. (ed), 1976: *Handbook of Industrial and organisational psychology*. Chicago: Rand McNally, pp. 363 - 413.

Edexcel, 2012: *GCSE Business Specification*. Accessed, 29/4/14 www.edexcel.com/quals/gcse/gcse09/Business/Business/Pages/default.aspx

Ellsworth, E. 1997: *Teaching Positions: Difference, pedagogy and the power of address.* Teachers College Press: New York

Entwistle, N. 1983: *Styles of Teaching and Learning: An integrated outline of educational psychology for students, teachers and lecturers*. Chichester: John Wiley & Sons Ltd.

Entwistle, N. 1987: *Understanding Classroom Learning*, London: Hodder and Stoughton.

Entwistle, N.1995: The use of research on student learning in quality assessment. In: G. Gibbs, (ed). *Improving Student Learning – Through Assessment and Evaluation*. Oxford: Oxford Centre for Staff Development (OCSD) pp. 24 - 43.

Entwistle, N. 2000: *Promoting Deep Learning Through Teaching and Assessment: Conceptual frameworks and educational contexts*. Accessed 6/1/05 <u>http://www.ed.ac.uk/etl/docs/entwistle2000.pdf</u>

Entwistle, N., 2003. *Concepts and Conceptual Frameworks Underpinning the ETL Project. Occasional Report 3*, Economic and Social Research Council, University of Edinburgh. Accessed 27/9/04 http://www.ed.ac.uk/etl/docs/ETLreport3.pdf

Entwistle, N. 2005: Conceptions of learning and the experience of understanding: Thresholds, contextual influences, and knowledge objects. Draft chapter for S. Vosniadou and A. Baltas, (eds). *Philosophical, Historical and Psychological Approaches to Conceptual Change* [in press]. (Personal communication 15th February 2005)

Entwistle, N., McCune, V. and Hounsell, J. 2002: Approaches to Studying and Perceptions of University Teaching-learning Environments: Concepts, measures and preliminary findings. Occasional Report 1. Accessed 17/02/05 http://www.ed.ac.uk/etl/publications.html

EORM, 2006: *Exploring Online Research Methods*. ReStore @ National Centre for Research Methods. Accessed 15/5/14 www.restore.ac.uk/orm/guestionnaire/guesads.htm

Ernest, P. 1994: Constructing Mathematical Knowledge: Epistemology and mathematics education. London: Falmer.

Field, A. 2009: Discovering Statistics Using SPSS 3e. London: Sage.

Flanagan, M. 2013: Threshold Concepts: Undergraduate teaching,

postgraduate training and professional development: A short introduction and bibliography. Accessed 5/7/14

http://www.ee.ucl.ac.uk/~mflanaga/thresholds.html

Fleming, C. and Bowden, M. 2009: Web-based surveys as an alternative to traditional mail methods, *Journal of Environmental Management*, **90**: pp. 284-292.

Fodor, J. 1998: When is a dog not a DOG? Nature, 396: pp. 325-327

Golafshani, N. 2003: Understanding reliability and validity in qualitative research. *The Qualitative Report*, **8** (4): pp. 597-607. Accessed 21/9/14 http://www.nova.edu/ssss/QR/QR8-4/golafshani.pdf

Friedrichs, J. and Kratochwil, F. 2009: On acting and knowing: How pragmatism can advance international relations research and methodology. *International Organisation*, **63**: pp. 701 – 731.

Graduate Student Instructor Teaching and Resource Center, University of California Berkeley. 2014: *Learning: Theory and research*. Accessed 8/2/14

http://gsi.berkeley.edu/gsi-guide-contents/learning-theory-research/

Gravetter, F.J. and Wallnau, L.B. 2011: *Essentials of Statistics for the Behavioural Sciences 7e.* Belmont, California: Wadsworth, Cengage Learning.

Griffiths, M. 1998: *Educational Research for Social Justice: Getting off the fence*. Buckingham: Open University Press.

Guba, E.G. 1981: Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Resources Information Center Annual Review*, **29**: pp. 75-91.

Guba, E.G. and Lincoln, Y.S. 2001: *Guidelines and Checklist for Constructivist* (*a.k.a. fourth generation*) *Evaluation*. Accessed 11/10/14

http://www.wmich.edu/evalctr/archive_checklists/constructivisteval.pdf

Haney, W., Russell, M., Gulek, C. and Fierros, E. 1998: Drawing on education: Using student drawings to promote middle school improvement. *Schools in the* Middle, 7 (3): pp. 38-43.

Harlen, W. 2007: Assessment of Learning. London: Sage.

Hattie, J.A.C. and Brown, G.T.L. 2004: *Cognitive Processes in asTTIe: The SOLO taxonomy.* asTTIe Technical Report #43: University of Auckland/Ministry of Education.

Heywood, J. 2000: Assessment in Higher Education: Student learning, teaching, programmes and institutions, 3rd revised edition. London: Jessica Kingsley Publishers.

Hilgard, E.R. and Bower, G.H. 1975: *Theories of Learning*. Englewood Cliffs, New Jersey: Prentice Hall.

Hoepfl, M.C. 1997: Choosing qualitative research: A primer for technology education researchers. *Journal of Technology Education*, **9** (1): pp. 47-63. Accessed 21/9/14

http://scholar.lib.vt.edu/ejournals/JTE/v9n1/hoepfl.html

Holsti, O.R. 1969: *Content Analysis for the Social Sciences and Humanities*. Reading, Massachusetts: Addison-Wesley.

Honey, P., and Mumford, A. 1986: *The Manual of Learning Styles.* Maidenhead, Berkshire: Peter Honey.

Hook, P. 2012: Teaching and learning: Tales from the ampersand. In L. Rowan and C. Bigum (eds), *Future Proofing Education: Transformative approaches to new technologies and student diversity in futures oriented classrooms.* Springer, pp. 123 and 124.

Hook, P. 2013: *Advantages of SOLO Taxonomy*. 25th November 2013. Accessed 30/01/15. <u>http://pamhook.com/wiki/Advantages_of_SOLO_Taxonomy</u>

Illeris, K. 2009: *Contemporary Theories of Learning: Learning theorists...in their own words*. Abingdon, Oxon: Routledge.

Jarvis, P., Holford, J. and Griffin, C. 2003: *The Theory and Practice of Learning 2e.* London: Kogan Page.

Jephcote, M. and Abbott, I. 2005: *Teaching Business Education 14-19*. London: David Fulton

Johnson, R.B. and Onwuegbuzie, A.J. 2004: Mixed methods research: A research paradigm whose time has come. *Educational Researcher.* **33**: pp. 14 - 26.

Johnson, R.B., Onwuegbuzie, A. J. and Turner, L.A. 2007: Toward a definition of mixed methods research. *Journal of Mixed Methods Research*. **1** (2): pp. 112 - 113

Jonassen, D.H. and Land, S.M. 2000: *Theoretical Foundations of Learning Environments*. Mahwah, New Jersey: Lawrence Erlbaum.

Jones, M. 2001: The issue of bias and positionality in cross-cultural, educational studies - enhancing the validity of data through a reflectivereflexive approach. Paper presented at the *Higher Education Close Up Conference 2*, Lancaster University, 16-18 July 2001. Accessed 25/1/14

http://www.leeds.ac.uk/educol/documents/00001733.htm

Joppe, M. 2000: The Research Process. Accessed 21/9/14

https://www.uoguelph.ca/hftm/research-process

Kahneman, D. 2011: *Thinking Fast and Slow*. New York: Farrar, Strauss and Giroux.

Kalat, J.W. 2005: *Introduction to Psychology 7e*. Belmont, California: Wadsworth, Thomson Learning.

Kelly, G. A. 1991: The Psychology of Personal Constructs: A theory of personality (Vol 1). London: Routledge.

Killen, R. 2005: *Programming and Assessment for Quality Teaching and Learning.* South Melbourne, Australia: Thomson Social Science Press.

Kirkman, T.W. 1996: Statistics to Use. Accessed 28/8/12

http://www.physics.csbsju.edu/stats/

Köhler, W. 1925: The Mentality of Apes. London: Kegan Paul, Trench and

Trubner.

Kolb, D.A. 1984: Experiential Learning: Experience as the source of learning and development. New Jersey: Prentice Hall.

Kralik, D. Price, K., Warren, J. and Koch, T. 2006: Issues in data generation using email group conversations for nursing research. *Journal of Advanced Nursing.* **53** (2): pp. 213 – 220.

Krefting, L. 1991: Rigor in qualitative research: The assessment of trustworthiness. *The American Journal of Occupational Therapy.* **45** (3): pp. 214-222.

Krippendorff, K. 2013: *Content Analysis: An introduction to its methodology*. Thousand Oaks, California: Sage.

Land, R. and Meyer, J.H.F. 2010: Threshold concepts and troublesome knowledge: Dynamics of assessment. In J.H.F. Meyer, R. Land, and C. Baillie, (eds.), *Threshold Concepts and Transformational Learning.* Rotterdam, Netherlands: Sense Publishers, pp.61-79.

Land, R., Meyer, J.H.F. and Baillie, C. 2010: Editors' Preface: Threshold Concepts and Transformational Learning. In J.H.F. Meyer, R. Land, and C. Baillie, (eds.), *Threshold Concepts and Transformational Learning.* Rotterdam, Netherlands: Sense Publishers, pp.ix-xlii.

Land, R. 2014: Liminality Close-up. Accessed 5/7/14

http://www.lancaster.ac.uk/fass/events/hecu7/docs/ThinkPieces/land.pdf

Laurillard, D. 2002: *Rethinking University Teaching: A framework for the effective use of learning technologies 2e.* Abingdon, Oxon: Routledge Falmer

Lefrançois, G.R. 2012: *Theories of Human Learning: What the professor said 6e*. Mason Ohio: Cengage Learning.

Lincoln, Y. S. and Gubba, E. G. 1985: *Naturalistic Inquiry.* Beverly Hills, California: Sage.

Lister, R., Simon, B., Thompson, E., Whalley, J. L. and Prasad, C. 2006: Not seeing the forest for the trees: Novice programmers and the SOLO taxonomy. Page | 286 Proceedings of the 11th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education. ACM: New York.

Mackenzie, N. and Knipe, S. 2006: Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research*, **16** (2): pp.193-205.

Madge, C. and O'Connor, H. 2002: Online with e-mums: Exploring the Internet as a medium for research, *Area*, **34** (1): pp. 92-102.

Mankiw, N. G., Taylor, M. P. and Ashwin, A. K. (contributory author), 2014: *Economics 3e.* Andover, Hampshire: Cengage Learning.

Martin, D.J. and Loomis, K.S. 2014: *Building Teachers: A constructivist approach to introducing education*. Mason Ohio: Cengage Learning.

Martin, S. 2011: Using SOLO as a Framework for Teaching: A case study in *maximising achievement*. Laughton, Sussex: Essential Resources Educational Publishers Limited.

Martin-Hays, J. 2008: Threshold and transformation. *European Journal of Management*, **8** (3): pp. 24-46.

Marton, F. 1981: Phenomenography: Describing conceptions of the world around us'. *Instructional Science*, **10**: pp. 177-200.

Marton, F. 1983: Beyond Individual Differences. *Educational Psychology*, **3**, (3) and (4): pp. 289-303.

Marton, F., Dall'Alba, G. and Beaty, E. 1993: Conceptions of Learning. International Journal of Educational Research, **19**: pp. 277-300.

Marton, F. and Säljö, R. 1976: On qualitative differences in learning: I – outcome and process. *British Journal of Educational Psychology*, **46**: pp. 4– 11.

Marton, F. and Säljö, R. 1984: Approaches to learning. In F. Marton, D. Hounsell, and N.J. Entwistle, (eds.), *The Experience of Learning*. Edinburgh:

Marzano, R.J. and Kendall, J. S. 2007: The New Taxonomy of Educational

Scottish Academic Press, pp. 39-58.

Objectives 2e. Thousand Oaks, California: Corwin Press.

McCune, V. 1998: Academic development during the first year at University. In C. Rust (ed.) *Improving Student Learning: Improving students as learners.* Oxford: Oxford Centre for Staff and Learning Development (OCSLD).

McCune, V. and Entwistle, N. 2000: *The Deep Approach to Learning: Analytic abstraction and idiosyncratic development*. Accessed 6/1/05

http://www.ed.ac.uk/etl/docs/mccune2000.pdf

McDowell, M., Thom, R., Pastine, I., Frank, R. and Bernanke, B. 2012: *Principles of Economics, 3rd European Edition.* Maidenhead, Berkshire: McGraw Hill.

Mertens, D.M. and McLaughlin, J.A. 2003: *Research and Evaluation Methods in Special Education*. Thousand Oaks, California: Corwin Press.

Mertens, D.M. 2015: *Research and Evaluation in Education and Psychology: Integrating diversity with quantitative, qualitative and mixed methods.* Thousand Oaks, California: Sage.

Meyer, J.H.F. 1995: A Quantitative Exploration of Conceptions of Learning. *Research and Development in Higher Education*, **18**: pp. 545-550

Meyer, J.H.F and Land, R. 2003: Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines. In C. Rust, (ed) *Improving Student Learning – Ten Years On.* Oxford: OCSLD, pp. 412-424.

Meyer, J.H.F and Land, R. 2005: Threshold concepts and troublesome knowledge 2: Epistemological considerations and a conceptual framework for teaching and learning. *Higher Education*, **49**: pp. 373-388.

Morgan, D.L. 2007: Paradigms lost and paradigms regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research.* **1** (1): pp. 48 - 76.

Morrow, S. L. 2005: Quality and trustworthiness in qualitative research in counselling psychology. *Journal of Counselling Psychology*, **52** (2): pp. 250-
260.

Morton, A. 1980: Frames of Mind. Oxford: Oxford University Press.

Moss, P, A., Girard, B J. and Haniford, L. C. 2006: Validity in educational assessment. *Review of Research in Education*, **30**: pp.109-162

Murray Thomas, R. 2000: *Comparing Theories of Child Development 5e.* Belmont, California: Wadsworth/Thomson Learning.

Näsström, G. 2009: Interpretation of standards with Bloom's revised taxonomy: A comparison of teachers and assessment experts. *International Journal of Research and Method in Education*. **32**(1): pp. 39-51.

Nolen-Hoeksema, S., Fredrickson, B., Loftus, G. and Wagenaar, W.A. 2009: *Atkinson and Hilgard's Introduction to Psychology 16e*. Andover, Hampshire: Cengage Learning.

Ofqual. 2011a: GCSE AS and A-level Subject Criteria for Economics. Ofqual/11/4957. Accessed, 15/7/14

http://ofqual.gov.uk/.../gce-as-and-a-level-subject-criteria-for-economics/

Ofqual. 2011b: GCSE Subject Criteria for Business Subjects. Ofqual/11/5088. Accessed, 15/7/14

http://ofqual.gov.uk/documents/gcse-subject-criteria-for-business-subjects/

Ofqual. 2011c: GCSE AS and A-level Subject Criteria for Business Subjects. Ofqual/11/4959. Accessed, 15/7/14

https://www.gov.uk/government/uploads/system/uploads/attachment_data /file/371158/11-10-18-gce-business-studies.pdf

Ofqual. 2011d: *GCSE*, *GCE*, *Principal Learning and Project Code of Practice*. May 2011. Accessed, 30/1/15

https://www.gov.uk/government/uploads/system/uploads/attachment_data /file/371268/2011-05-27-code-of-practice.pdf

Ofqual. 2013: Standards. 26th April, 2013. Accessed, 29/4/14.

www.ofqual.gov.uk/standards.

Ofqual. 2014a: Validity. 11th February 2014. Accessed, 3/5/14.

www.ofqual.gov.uk/standards/research/validity

Ofqual. 2014b: Reliability. 9th October 2013. Accessed, 3/5/14.

www.ofqual.gov.uk/standards/research/reliability

Ofqual. 2014c: Review of Quality of Marking in Exams in A-levels, GCSEs and Other Academic Qualifications. Final Report. Ofqual/14/5379. Accessed, 16/3/14

http://ofqual.gov.uk/documents/quality-of-marking-final-report/

Ofqual. 2014d: *Statistical Release*. Accessed 27/10/14 <u>http://ofqual.gov.uk/standards/statistics/enquiries-about-results/</u>

Ofsted. 2002: *The Key Stage 3 Strategy: Evaluation of the first year of the pilot*. London: Office for Standards in Education (Ofsted), HMI 349. Accessed 23/12/13

http://dera.ioe.ac.uk/16531/1/Key%20Stage%203%20strategy.pdf

Ofsted. 2003: *Good Assessment Practice in Business Studies*. London: Office for Standards in Education (Ofsted), HMI 1471.

Ofsted. 2008: Developing Young People's Economic and Business Understanding: Business education in secondary schools, colleges and initial teacher training 2004/07. Reference number 070096. London: Office for Standards in Education.

Ofsted. 2011: *Economics, Business and Enterprise Education: A summary of inspection evidence, April 2007 – March 2010.* Reference no: 10086. Accessed 6/12/14

https://www.ofsted.gov.uk/sites/default/files/documents/surveys-and-goodpractice/e/Economics%20business%20and%20enterprise%20education.pdf

Olsen, M.H. and Hergenhahn, B.R. 2009: *An Introduction to Theories of Learning 8e.* Upper Saddle River, New Jersey: Pearson.

Parkin, M., Powell, M. and Matthews, K. 2008: *Economics 7e.* Harlow, Essex:

Pearson Education.

Patton, M.Q. 2002: *Qualitative Research and Evaluation Methods 3e.* Thousand Oaks, California: Sage.

Peel, E.A. 1971: *The Nature of Adolescent Judgment*. London: Staples Press Ltd.

Perkins, D. 1999: The constructivist classroom: The many faces of constructivism. *Educational Leadership*. **57** (3): pp. 6-11.

Perkins, D. 2006: Constructivism and troublesome knowledge. In J.H.F. Meyer, and R Land, (eds). *Overcoming Barriers to Human Understanding: Threshold concepts and troublesome knowledge*. London: Routledge, pp. 33-47.

Pinker, S. 2002: The Blank Slate. London: Penguin

Polanyi, M. 1958: *Personal Knowledge: Towards a post-critical philosophy*. London: Routledge and Keegan Paul.

Pollitt, A. and Ahmed, A. 2008: *Outcome Space Control and Assessment*. Paper prepared for the 9th Annual Conference of the Association for Educational Assessment, Hissar, Bulgaria, 6th - 8th November, 2008. Accessed 25/10/14

http://www.lifeinbits.org/camexam/htdocs/papers/OutComeSpaceAPAA.pdf

Prosser, M. and Trigwell, K. 1997: 'Relations between perceptions of the teaching environment and approaches to teaching', *British Journal of Educational Psychology*, **67**: pp. 25-35.

Prosser, M. and Trigwell, K. 1999: Understanding Learning and Teaching. The experience of higher education. Buckingham: Open University Press.

Prosser, M., Trigwell, K., Hazel, E. and Waterhouse, F. 2000: Students' experiences of studying physics concepts: The effects of disintegrated perceptions and approaches. *European Journal of Psychology of Education*. XV: pp. 61-74. Purdy, J.E., Markham, M.R., Schwartz, B.L. and Gordon, W.C. 2001: *Learning and Memory 2e.* Belmont, California: Wadsworth/Thomson Learning.

Qingping, H. and Opposs, D. 2012: The reliability of results from national tests, public examinations and vocational qualifications in England. Educational Research and Evaluation: An international journal on theory and practice. **18** (8): pp. 779-799.

Quinlan, C. 2011: *Business Research Methods*. Andover, Hampshire: Cengage Learning.

Ramsden, P. 1992: *Learning to Teach in Higher Education*, London: Routledge.

Rayner, S. G. 2013: Problematising style differences theory and professional learning in educational psychology. *The Australian Educational and Developmental Psychologist.* **30** (1): pp. 13 – 51.

Rayner, S. 2014: Personal correspondence. 20/3/14.

Riessman, C.K. 2008: *Narrative Methods for the Human Sciences*. Thousand Oaks, California: Sage.

Robson, C. 2002: Real World Research 2e. Oxford: Blackwell

Rosch, E. 1978: Principles of Categorisation in E. Rosch and B.B. Lloyd (eds). *Cognition and Categorisation*. Hittsdale, New Jersey: Erlbaum, pp. 27-48.

Ross, L. and Anderson, C.A. 2008: Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic and A. Tversky, (eds), *Judgement Under Uncertainty: Heuristics and biases*. New York: Cambridge University Press, pp. 129-152.

Rowbottom, D. P. 2007: Demystifying threshold concepts. In *Journal of Philosophy of Education*, **41** (2): pp. 263-270.

Rubin, P.H. 2002: *Folk Economics*. 19th September 2002. Accessed 25/8/05 2005. <u>http://ssrn.com/abstract=320940</u>

Rust, C. 2006: *Improving Student Learning: Improving student learning through assessment*. Oxford: Oxford Centre for Staff and Learning

Page | 292

Development (OCSLD).

Sambell, K., McDowell, L and Brown, S. 1997: "But is it fair?": an exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, **23** (4), pp. 349-371

Sanneke, B. and Voeten, J. M. 2004: Teachers' conceptions of student learning and own learning. *Teachers and Teaching: Theory and practice*, **10** (1): pp. 77-98.

Scottish Qualifications Authority. 2014: *Guide to Assessment*. Accessed, 21/4/14 www.sqa.org.uk/files_ccc/25GuideToAssessment.pdf

Shaikh, A. 2013: On the role of reflexivity in economic analysis. *Journal of Economic Methodology*, **20**, (4): pp. 439-445.

Shenton, A. K. 2004: Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, **22**: pp. 63-75.

Skinner, B. F. 1948: Superstition in the pigeon. *Journal of Experimental Psychology*, **38**: pp. 168-172.

Skinner, B.F. 1950: Are theories of learning necessary? *Psychological Review*, **57**, (4): pp. 193-216

Soros, G. 2013: Fallibility, reflexivity and the human uncertainty principle. *Journal of Economic Methodology.* **20** (4): pp. 309–329.

Spender, J.C. 1996: Organizational knowledge, learning and memory: three concepts in search of a theory. *Journal of Organizational Change Management*, **9**, (1): pp. 63-78.

Stemler, S. 2001: An overview of content analysis. *Practical Assessment*, *Research and Evaluation*, **7** (17): pp. 137-146.

Stenbacka, C. 2001: Qualitative research requires quality concepts of its own. *Management Decision*, **39** (7): pp. 551-555.

Stenhouse, L. 1975: *An Introduction to Curriculum Research and Development*. London: Heinemann.

Sternberg, R.J. 2003: *Cognitive Psychology 3e.* Belmont, California: Wadsworth/Thomson Learning.

Strathclyde, 2014: Perspective, Positionality, Bias and Reflexivity. Accessed 14/6/14

http://www.strath.ac.uk/aer/materials/1educationalresearchandenquiry/uni t4/perspectivepositionalitybiasandreflexivity/

Tashakkori, A. and Teddlie, C. (eds) 2003: *Handbook of Mixed Methods in Social and Behavioural Research*. Thousand Oaks, California: Sage.

Tashakkori, A. and Creswell, J.W. 2007: Exploring the nature of research questions in mixed methods research. *Journal of Mixed Methods Research*. **1** (3): pp. 207 – 11

Teddlie, C. and Tashakkori, A. 2009: Foundations of Mixed Methods Research: Integrating quantitative and qualitative approaches in the social and behavioural sciences. Thousand Oaks, California: Sage.

Thorndike, E.L. 1923: *Education Psychology: Briefer course*. Abingdon, Oxon: Routledge

Times Educational Supplement. 2013: *The Perfect (OFSTED) Lesson Plan.* Accessed, 23/12/13

http://www.tes.co.uk/teaching-resource/The-Perfect-OFSTED-lesson-plan-6115887/

Tversky, A. and Kahneman, D. 2008: Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic and A. Tversky, (eds), *Judgement Under Uncertainty: Heuristics and biases*. New York: Cambridge University Press, pp. 3 - 20.

Tyler, R.W. 1949: *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.

University of California, Berkeley. 2011: *Teaching Guide for Graduate Student Instructors. Learning: Theory and research.* Accessed, 7/6/14.

http://gsi.berkeley.edu/teachingguide/theories/index.html

Vivian, P. 2012: A new symbol based writing system for use in illustrating basic dynamics. Unpublished PhD thesis. Coventry University. Accessed, 15/7/14

http://www.heacademy.ac.uk/assets/documents/subjects/engineering/EE20 10/45_RDP_Vivian.pdf

von Glasersfeld, E. 1994: A radical constructivist view of basic mathematical concepts. In, P. Ernest, (ed). *Constructing Mathematical Knowledge: Epistemology and Mathematics Education*. London: Falmer, pp. 5-7

Vygotsky, L.S. 1978: *Mind and Society: The development of higher mental processes*. Cambridge, Massachusetts: Harvard University Press.

Walker, G. 2013: A cognitive approach to threshold concepts. *Higher Education*, **65**: pp. 247-263

Watkins, D. 1983: Depth of processing and the quality of learning outcomes. *Instructional Science*, **12**: pp. 49 – 58.

Weber, R. P. 1990: Basic Content Analysis 2e. Newbury Park, California: Sage.

Weiskopf, D. A. 2014: *The Theory-Theory of Concepts*. Internet Encyclopedia of Philosophy. Accessed 26/6/14 <u>http://www.iep.utm.edu/th-th-co/</u>

Wenger, E. 2006: *Communities of Practice: Learning, meaning and identity.* Cambridge: Cambridge University Press.

Wiliam, D. 2000: The meanings and consequences of educational assessments. *Critical Quarterly*, **42** (1): pp. 105 – 127.

Wittgenstein, L. 1968: *Philosophical Investigations*. Translated by G.E.M. Anscome. Oxford: Basil Blackwell.

Wood Wallace, D. 2012: *Is Reflexivity a Sufficient Protection Against the Risks of Positionality?* Unpublished Essay (Masters). Nottingham Trent University. Accessed 27/1/14

http://www.academia.edu/3365173/_2012_ls_reflexivity_a_sufficient_protec tion_against_the_risks_of_positionality Wragg, E. C. 2001: Assessment and Learning in the Secondary School. London: Routledge Falmer.

Yarnold J. K. 1970: The minimum expectation in chi-square goodness of fit tests and the accuracy of approximations for the null distribution. *Journal of the American Statistical Association*, **65** (330): pp. 864-886

Zalta, E. N. 2012: *Gottlob Frege.* Stanford Encyclopedia of Philosophy. Accessed, 28/6/14 <u>http://plato.stanford.edu/entries/frege/#FreTheSenDen</u>

9. Appendices

Appendix I

Grade descriptions from Edexcel Economics and Business Specification.

Grade A

Candidates recall accurately, select and communicate their detailed knowledge and thorough understanding of concepts, theories, issues and current practice of business. They understand and use business terminology accurately and appropriately.

They plan and carry out independently and effectively a range of investigations and tasks using a wide range of skills competently, making effective use of time and resources. They apply their knowledge and critical understanding effectively to select and organise information from a wide range of sources and to investigate business organisations in a variety of contexts.

They use and evaluate quantitative and qualitative evidence effectively with a high degree of accuracy to analyse problems and issues, and make informed and reasoned judgements to present conclusions accurately and appropriately.

Grade C

Candidates recall, select and communicate sound knowledge and understanding of concepts, theories, issues and current practice in business. They use business terminology appropriately.

They plan and carry out different investigations and tasks using a range of skills appropriately. They apply their knowledge and understanding to select relevantly and organise information from a variety of sources to investigate business organisations in different contexts.

They use and evaluate evidence to analyse problems and issues with some accuracy, make reasoned judgements and present conclusions that are supported by evidence.

Grade F

Candidates demonstrate knowledge and understanding of some basic aspects of concepts, theories, issues and current practice in business. They use some simple business terminology.

They carry out straightforward investigations and tasks using a limited range of skills. They apply some knowledge and understanding to select and organise simple information from a limited range of sources to investigate business organisations.

They identify simple evidence to analyse problems and issues and make judgements. They present simple conclusions that are sometimes supported by evidence.

Appendix II

E-mail text to prospective participants outlining the task for Phase II. Dear

I am doing a PhD and the research I am conducting is centred on assessment at GCSE and how students learn and understand economics and business at this level. This is the second phase of field work in the research which involves looking at ways in which different assessment taxonomies might be used by teachers.

The work involves marking some student work - about 1500 words in total. The work is a series of answers by students to GCSE examination questions. I want volunteers to mark these pieces of work. You will be provided with a mark scheme for doing so. Initially you will be sent the answers with a mark scheme and asked to mark the work in accordance with that mark scheme. Six weeks later you will be sent another mark scheme and asked to mark the same work according to that mark scheme. That's it! I would anticipate that the actual marking would take no more than two hours at most and maybe half an hour to familiarise yourself with the mark scheme.

Those taking part will be entered into a draw for three prizes; first prize is a copy of the Active Teach Assessor CD ROM for the new Business Studies GCSE from Edexcel worth £300, second prize is a copy of the Introduction to Small Business textbook with Teacher's Guide and third prize will be a copy of the new Controlled Assessment Handbook for Students.

If you are interested in pursuing this further, I need to 'recruit' 20 people for the task. I will send out full details of the mark scheme and the task and the research permissions forms (required as part of the ethical guidelines of the University I am based at) once you have indicated whether you are interested in taking part. If you are not interested or not able to take part because of time commitments then I quite understand.

Regards

Andrew

Appendix III

Permissions letters for Phases II and III of the Research

Phase II

Explanatory Introduction and Permissions Form

PhD Research Assessment Task

Introduction:

Firstly, thank you for agreeing to take part in this research. Your cooperation is greatly appreciated.

You will need to have some experience of teaching economics and business to the relevant age group in this study (aged 14 - 16 and studying GCSE or Level 2 qualifications).

The PhD has the following title:

Exploring the problematic nature of GCSE examining in Economics and Business: Assessing troublesome knowledge, threshold concepts and learning.

I am researching the idea of threshold concepts, troublesome knowledge and assessment of learning in economics and business subjects. A threshold concept acts as a 'portal'; once a student enters the portal they will realise a transformed understanding of the subject - they will 'think like an economist.' i.e. they begin to approach problems, questions and issues in a new way. Many students experience difficulty with threshold concepts because they are often counter intuitive. Prior to passing through the 'portal' students will experience troublesome knowledge. Learning at this stage remains at a surface level characterised by relying on rote memorisation to make progress. Students will frequently experience difficulties in transferring understanding of these concepts to new contexts.

I am making an informed assumption that a transformed understanding of the subject is a desirable learning outcome as opposed to surface learning. I am researching how students might be helped to reach this transformed understanding and the assessment regime is part of the jigsaw. Part of the role of the teacher is in assessing student learning. The assessment models in use today are invariably based on Bloom's Taxonomy. The research will look at whether this is the best model to use if we want to help students 'think like an economist'.

This part of the research, which you have agreed to take part in, presents a series of answers to examination questions by students who have been studying economics and business as part of a GCSE qualification.

This part of the research is based on two examination questions; there are eight separate responses to question 1 (total word count 785) and five to Question 2 (total word count 730).

Each answer has been reproduced via word processing to make it easy to read but the language, spelling, punctuation and so on, are all faithfully reproduced.

You will be given these answers to mark.

You will be given a mark scheme based on Bloom's Taxonomy and a mark scheme based on an alternative taxonomy called the SOLO Taxonomy.

You will be required to mark the same answers using these two taxonomies. There will be a time period of 6 weeks between receiving the first taxonomy and returning your marks and the second. Both mark schemes will contain further information about the taxonomy and how to mark the answers.

You will receive the mark schemes and the answers electronically and will be asked to return both by the same method. Comments/observations can be made on the answers using the 'Insert comment' feature of your word processing software or simply by typing in your comments in a different colour.

The research write up will not identify you personally but refer to you as 'Assessor 1', 'Assessor 2' etc. Prior to receiving the first set of answers, I require you to read the following form. If you understand and agree to the conditions then please complete the personal details and return them to me via e-mail. An electronic signature or typed signature is acceptable.

Thank you again for agreeing to help.

Mallez

PhD Research: Permission Form Researcher - Andrew Ashwin

By signing the form, you agree:

- That the researcher can use the information generated and any comments you make for the purposes of this research.
- The information generated will be used exclusively as part of the research for a PhD at the awarding University.
- Parts of this research or the whole research may be published in academic journals or other such publications.
- Extracts from your comments may be used to illustrate particular points in the thesis. This thesis, or work derived from this thesis, may be published and therefore appear in the public domain.
- Your comments will remain anonymous at all times.
- PLEASE NOTE you will NOT be paid for the contribution you have agreed to make.
- The details below will help the researcher to be able to contact you should a query arise about the use of your work in the research. You do not have to provide personal details if you do not wish to do so.
- Your personal details will not be shared with any other person, institution or organisation, apart from the researcher's supervisors and/or examiners, without your expressed permission.

If you agree to these terms please check the box below.

 $\hfill\square$ I have read and understand the conditions above and agree to take part in this research.

Signature
Date:
Name you wish to be known as - PLEASE PRINT:
E-mail address:
Address (optional):
GCSE Subject/s taught and awarding body:

Phase III

Explanatory Introduction and Permissions Form

PhD Research

Dear Colleague

The questions for this phase of research are designed to be answered via email exchange. Respondents should open the Word document, save it with a suitable file name and complete the questions by typing in your answer to each question under the question. There are 12 questions. The space under each question will expand as you type.

The PhD has the following title:

Exploring the problematic nature of GCSE examining in Economics and Business: Assessing troublesome knowledge, threshold concepts and learning.

The research focuses on GCSE level and as a result it is preferable that you answer the questions with this level of study and student in mind. The research will look at the subjects of economics and business. Please specify what GCSE subject/s you teach in the permission form below. This might include Business Studies, Economics or a combined Economics and Business course.

Please read the permissions form <u>before</u> you begin the answers to the questions and complete the contact details where appropriate. If there is any part of the permissions that you do not feel you can accept then please contact me.

I am very grateful for your offer of help in this phase of the research. Please will you complete the form below and return it to me along with your completed answers to the questions to:

xxxxxxxxx@btopenworld.com

Thank you again for your help.

Mallez

Andrew Ashwin PhD Research: Permission Form Newman University College The University of Leicester. Researcher - Andrew Ashwin

By signing the form, you agree:

- That the researcher can use your answers as part of a research programme for PhD study.
- The research will be used exclusively as part of the research for a PhD at Newman University College, accredited by the University of Leicester, and for no other purpose.
- Your comments will remain anonymous at all times.
- Extracts from your comments may be used to illustrate particular points in the thesis. This thesis, or work derived from this thesis, may be published and therefore appear in the public domain.
- PLEASE NOTE you will NOT be paid for the contribution you have agreed to make.
- You will NOT be able to withdraw your consent once you have signed this form.
- The details below will help the researcher to be able to contact you should a query arise about the use of your work in the research. You do not have to provide personal details if you do not wish to do so.

Full name - PLEASE PRINT:

Signature (an electronic	
signature is acceptable):	
Date:	
E-mail address:	
Address (optional):	
CCSE Subject/s taught and	

GCSE Subject/s taught and awarding body:

Questions

- 1. What is your definition of 'learning'?
- 2. What would you say are the top five economics and business concepts which students must learn?
- 3. How important is it for students to learn key terms in business and economics to tackle the examination effectively?
- 4. A student arrives at your lesson with an interesting news article which they have found on a topic that you covered in a lesson two weeks ago and asks a question about it. You have now moved on to a different topic. How, if at all, would you use the student's query in your lesson?
- 5. What do you think are the **two** main barriers to improved student understanding of business and economics at GCSE?
- 6. Ofqual has referred to 'thinking in the subject' thinking like an historian, thinking like an artist, thinking like an economist etc. where students have a transformed understanding of the subject.

How important do you think it is to make 'thinking in the subject' a key focus of your teaching and learning programme at GCSE level?

- 7. What <u>one</u> change to your current situation would improve the learning of your students at GCSE?
- 8. How important is it for you to know that you have covered the entire examination syllabus by the end of the academic year?
- 9. Do you think writing frames are a useful aid to help students learn economics and business? Please explain.
- 10. Thinking about exam questions in GCSE Business and Economics, are there question types that are not included on papers at present which you would like to see included and why?
- 11. How do you know when a student demonstrates learning in your lesson?
- 12. A student's written work demonstrates evidence of some confusion over a key economic/business concept. What plans do you put in place to help such a student?

Appendix IV

Spreadsheet logging replies for Phase II

Name	Date agreed	Permission Returned	Mark Scheme	MS1 Returned	Mark Scheme	MS2 Returned	Date to send second MS
1 Ass'r1	14/05/2010	17/05/2010	18/05/2010	24/05/2010	05/07/2010	23/08/2010	05/07/2010
2 Ass'r2	17/05/2010	17/05/2010	05/07/2010	24/07/2010	18/05/2010	24/05/2010	05/07/2010
3 Ass'r3	27/05/2010	28/05/2010	28/05/2010	10/06/2010	18/07/2010	21/09/2010	22/07/2010
4 Ass'r4	11/06/2010	11/06/2010	11/06/2010	30/07/2010	23/08/2010	31/08/2010	20/08/2010
5 Ass'r5	27/05/2010	07/06/2010	07/06/2010	09/06/2010	18/07/2010	21/07/2010	22/07/2010
6 Ass'r6	27/05/2010	07/06/2010	07/06/2010	09/06/2010	18/07/2010	21/07/2010	22/07/2010
7 Ass'r7	27/05/2010	15/06/2010	15/06/2010	05/07/2010	10/08/2010	23/08/2010	16/08/2010
8 Ass'r8	27/05/2010	28/05/2010	27/07/2010	12/08/2010	28/05/2010	18/06/2010	30/07/2010
9 Ass'r9	27/05/2010	03/06/2010	12/07/2010	14/07/2010	04/06/2010	15/06/2010	12/07/2010
10 Ass'r10	27/05/2010	28/05/2010	18/07/2010	24/07/2010	28/05/2010	10/06/2010	22/07/2010
11 Ass'r11	28/05/2010	10/06/2010	10/06/2010	25/06/2010	10/08/2010	02/09/2010	06/08/2010
12 Ass'r12	28/05/2010	01/06/2010	18/07/2010	23/09/2010	01/06/2010	08/06/2010	19/07/2010
13 Ass'r13	28/05/2010	07/06/2010	07/06/2010	03/07/2010	10/08/2010	11/08/2010	08/08/2010
14 Ass'r14	28/05/2010	07/06/2010	27/07/2010	24/09/2010	07/06/2010	18/06/2010	30/07/2010
15 Ass'r15	07/06/2010	07/06/2010	10/08/2010	23/09/2010	07/06/2010	23/06/2010	04/08/2010
16 Ass'r16	07/06/2010	07/06/2010	10/08/2010		07/06/2010	27/06/2010	10/08/2010
17 Ass'r17	07/06/2010	07/06/2010	07/06/2010	10/06/2010	18/07/2010	09/09/2010	22/07/2010
18 Ass'r18	09/06/2010	09/06/2010	10/08/2010	31/08/2010	10/06/2010	21/06/2010	02/08/2010
19 Ass'r19	11/06/2010	28/06/2010	28/06/2010	06/07/2011	10/08/2010	31/08/2010	17/08/2010
20 Ass'r20	17/06/2010	17/06/2010	18/07/2010	31/07/2010	17/06/2010	23/06/2010	28/07/2010

Appendix V

Phase I Text Analysis Inter-coder Reliability Pilot

The following are definitions I have given to a series of categories that I am seeking to identify in a sample of scripts from students over a number of years. Each of the scripts is from exemplar material used in standardisation so you can assume that they have been marked in accordance with the mark scheme at the time and are, therefore, 'accurate' marks.

The sample of scripts I am looking at all feature students who have scored high marks and would have been 'A' grade candidates or above. What I am looking at the scripts to try and identify are examples of where such students might have demonstrated a lack of understanding or 'thinking like an economist'. Such examples may, or may not, be present in the scripts.

Each of the categories below has an associated number. What I would like you to do on the 2 scripts I gave you is to highlight (literally – using a highlighter) and note the relevant code number if/when you identify the particular category in the script. Please use distinctive pen colour to indicate where you have identified any relevant code.

The categories, code number in each case and the definitions are:

1. Contradictory knowledge:

Where 2 or more pieces of knowledge are used which are linked by the student but where their use contradicts what would be expected to happen or in an incorrect way.

2. Conceptual misunderstanding:

Where a particular economic/business concept is used or is referred to which demonstrates misunderstanding of that concept.

3. Partial understanding/knowledge:

Where a student is using knowledge but where the answer suggests that the student has an incomplete or partial grasp of that knowledge.

4. Structured answer to higher tier questions:

Where a higher tariff question is answered in a way that includes sections covering advantages/benefits/pros, disadvantages/costs/cons and conclusion. (In this case you can write the code number at the end of the answer rather than highlighting).

5. No assessment/evaluation:

Where a student is required by the question to make some evaluative comment but where points are raised in the answer without any evaluation/assessment given.

6. Unsupported assumptions:

Where a student makes an assumption of cause and effect without any evidence being provided in support.

Thank you in advance for your help.

Andrew

Appendix VI

Pilot Introduction and questions used at EBEA Conference in anticipation of Phase III.

Dear Colleague

The following represents a pilot for questions that will be used at a series of focus groups later this year. The aim of this pilot is to get some idea and understanding of whether these questions will elicit the sort of response I am looking for and help to guide the discussion at the focus group in the appropriate way to help gather research information. The feedback I receive from you will help me to refine the questions to help improve the quality of the data I hope to gather from the focus groups.

The PhD has the following title:

Exploring the problematic nature of GCSE examining in Economics and Business: Assessing troublesome knowledge, threshold concepts and learning.

The research focuses on GCSE level and as a result it is necessary for you to answer questions with this level of study and student in mind. The research will look at the subjects of economics and business. You may teach Business Studies, Economics or a combined Economics and Business course; please specify this in the permission form below.

There are a number of key research questions that I am seeking to answer. These include:

What are teachers' conceptions of learning in Economics and Business at GCSE level?

What are threshold concepts in economics?

How do current assessment structures influence assessment outcome?

What is the relationship between deep learning and 'thinking like an economist' at GCSE Level?

What effect would an alternative assessment regime have on teaching and learning in Economics and Business at GCSE level?

Page | 311

I am very grateful for your offer of help in this pilot study. Please will you complete the form below to indicate your agreement to participate in this study?

If you have any questions or observations following the session you can contact me on:

xxxxxxxx@xxxxxxx.com

Appendix VII

Pilot Questions for Phase III

Questions:

- What are the 5 most important concepts that you have to teach students at GCSE (these do not have to be listed in any particular order)?
- 2. Describe your role as a teacher of business and economics at GCSE level.
- Provide a definition of what you believe is meant by the term 'learning' at GCSE level in this discipline.
- 4. QCA has referred to the idea of 'thinking like an economist'. What do YOU think would constitute thinking like an economist?
- 5. What do you think are the two main barriers to improved student understanding of business and economics at GCSE?
- 6. Briefly comment on your views of the validity and reliability of the scheme of assessment of economics and business at GCSE level.
- To what extent do you feel the current assessment objectives help promote learning and understanding of business and economics at GCSE?

Appendix VIII

Phase III Questions

- 1. What is your definition of 'learning'?
- 2. What would you say are the top five economics and business concepts which students must learn?
- 3. How important is it for students to learn key terms in business and economics to tackle the examination effectively?
- 4. A student arrives at your lesson with an interesting news article which they have found on a topic that you covered in a lesson two weeks ago and asks a question about it. You have now moved on to a different topic. How, if at all, would you use the student's query in your lesson?
- 5. What do you think are the **two** main barriers to improved student understanding of business and economics at GCSE?
- 6. Ofqual has referred to 'thinking in the subject' thinking like an historian, thinking like an artist, thinking like an economist etc. where students have a transformed understanding of the subject. How important do you think it is to make 'thinking in the subject' a key focus of your teaching and learning programme at GCSE level?
- 7. What <u>one</u> change to your current situation would improve the learning of your students at GCSE?
- 8. How important is it for you to know that you have covered the entire examination syllabus by the end of the academic year?
- 9. Do you think writing frames are a useful aid to help students learn economics and business? Please explain.
- 10. Thinking about exam questions in GCSE Business and Economics, are there question types that are not included on papers at present which you would like to see included and why?
- 11. How do you know when a student demonstrates learning in your lesson?

12. A student's written work demonstrates evidence of some confusion over a key economic/business concept. What plans do you put in place to help such a student?

Appendix IX

Bloom and SOLO mark scheme record sheets

Bloom Marks:

					-		-													
Assessor	1	2	3	4	5	6	/	8	9	10	11	12	13	14	15	16	17	18	19	20
Boston Matr	ix (marks	out of 10)																		
Student 1	3	8	6	6	6	10	8	7	8	9	5	7	8	7	7	8	8	8	6	9
Student 2	2	4	3	1	2	4	2	4	3	6	2	3	4	4	4	5	4	4	3	5
Student 3	2	1	5	6	3	4	3	1	3	3	1	0	2	2	1	6	2	2	1	0
Student 4	6	7	4	7	5	4	3	4	5	5	2	3	3	4	5	6	6	6	7	0
Student 5	2	1	2	1	2	5	1	1	2	4	1	1	1	1	4	5	3	1	4	0
Student 6	2	6	5	6	4	6	4	6	6	6	5	4	3	4	3	7	6	0	3	6
Student 7	1	0	0	0	4	2	0	0	0	1	3	0	0	0	1	3	0	0	0	0
Student 8	2	6	5	7	3	4	6	2	5	4	4	2	4	4	3	5	4	3	2	4
Marker																				
Olympic Ga	mes (marl	s out of 1	6)																	
Student 1	11	10	7	12	9	6	9	9	11	11	12	7	9	9	10	10	10	8	10	9
Student 2	12	12	8	12	13	6	8	6	8	8	8	8	7	10	6	8	8	8	8	7
Student 3	8	10	10	8	7	7	8	8	7	9	8	8	8	10	9	9	8	6	6	6
Student 4	13	13	11	6	8	9	10	11	12	9	12	11	13	13	11	9	12	9	12	9
Student 5	9	9	7	5	5	5	12	10	11	10	10	10	7	9	7	8	10	4	3	5

SOLO Marks

Assessor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Boston Matri:	x (level)																			
Student 1	5	4	4	3	4	3	4	4	4	4	5	4	4	3	4	4	4	3	5	5
Student 2	2	2	2	2	3	2	2	2	2	2	2	2	2or3	2	2	2	3	2	1	2
Student 3	2	2	2	2	2	3	3	1	2	1	1	2	1	1	1	3	2	1	1	1
Student 4	4	3	3	42	or 3	3	3	1	2	2	2	4	1	1	2	3	2	3	2	1
Student 5	2	1	4	11	or 2	4	2	1	2	1	1	3	1	1	2	3	1	1	2	1
Student 6	4	3	5	3	3	4	3	4	3	3	3	2	2	3	4	3	3	1	4	4
Student 7	1	1	1	1	1	2	1	1	0	1	1	0	0	1	1	2	0	1	0	1
Student 8	4	4	4	4	2	3	3	2	2	4	4	2	3	2	3	4	2	4	2	2
Olympic Gan	nes (level)																			
Student 1	4	3	3	3	4	3	4	3	3	2	4	2	3	4	2	4	3	2	3	4
Student 2	4	4	4	53	or 4	4	4	3	3	3	4	3	4	3	3	4	3	3	3	3
Student 3	3	4	3	3	4	4	3	3	3	3	4	4	3	3	2	3	3	2	4	3
Student 4	4	5	4	4	5	4	3	4	4	2	4	4	4	4	1	4	4	4	5	4
Student 5	3	4	350	or 4	4	4	4	3	3	4	4	5	5	3	3	3	3	4	1	3

Appendix X

RESC Submission Form

Andrew Ashwin

Student Number s0712293

Supervisors: Professor Steve Rayner, Professor Mary Fuller.

Summary of issue to be considered by representatives of the ethics committee, 5th June 2008. 9.00am.

PhD title:

Exploring the problematic nature of GCSE Examination in Economics and Business: Assessing troublesome knowledge, threshold concepts and learning.

One element of the research involves textual analysis of student responses to exam papers. The data is in the possession of the researcher through his position as Chief Examiner of Economics and Business (GCSE) for Edexcel, the awarding body. The exam scripts are papers that have been used as exemplar material for standardisation meetings in examination series dating back to 2003.

I applied to Edexcel for permission to use the scripts for my research. Edexcel, as a commercial organisation owned by Pearson Education, has previously used exemplar scripts for teacher training and as part of the content of the Examiner's Report published for use by teachers. However, Edexcel are now pointing out that it does not own the IPR of the scripts and that the scripts are owned by the individual student. Since Edexcel do not own the IPR, it is not in a position to grant permission for their use in research. Edexcel has said that I can use the data but at my own risk.

Getting the free and informed consent of the individual students from the exemplar scripts would be impossible. For one, Edexcel would have to release the names of the students and centre (school or college) and it would not be in a position to do this even if it had the information going back as far as 2003.

In addition since 2005 scripts have been subject to being anonymised to prevent any suggestion of bias in the marking process. For scripts from 2005, therefore, the researcher does not have access to a name or to the centre from which the student was entered.

In using this data, I have considered the extent of the risk to myself as a researcher, and to the University. The risk is that a student reading the thesis, or any published works that arise out of the research, might be in a position to identify the work as his or her own and in so doing take exception to the fact their work has been used without their free and informed consent.

The risk, however, is negligible. I will negate any remaining risk in the following ways:

- Extracts only will be used not complete answers or scripts.
- The researcher does not possess information on who the students are or from what centre they were from.

Appendix XI

Phase I category spreadsheet

Student 1	Question	Researcher	Marker 1	Marker 2	Marker 3		Agreed
	1ai			3			
	1aii	3,3	3,3	6			4
	1b	6,6	6	6,6			5
	1c	5	3,5	3			2
	1d	3,3,5,6,6,6	3,3,3,5	5	5		7
	3ai						
	3aii	2,2,6	2,2,6	3	2,3		7
	3aiii	2	1,2				2
	3bi	6					
	3bii	2,2,3,6	2,3,6,6	2,2	3,6		10
	3biii	1,2,3,3,4,6,6,6,6	1,1,1,2,2,3,4	1,3,6,6	4,6,6		16
		29	25	13	8	75	
Student 2	1a	2	2,3	2	2,2		5
	1bi	3	3	3			3
	1bii						
	1ci	6,6	3	6	4		2
	1cii	3,5	3,5	3,6	3		6
	1di	3	3	1	3		3
	1dii						
	1e	4,6	1,4,	6,6	4,4		5
	2ai						
	2aii	2,5	5	2	2		5
	2b						
	2c	3	5	3			2
	2di	3	3	3	2		3
	2dii						
	2e	3,6	1,1,2,3,6	4,6	4		6
		18	17	13	6	54	
						129	93
							72.09

Appendix XII

Explanation of Bloom Mark Scheme

Assessment taxonomy based on Bloom:

This assessment regime is based on Bloom's Taxonomy, which forms the basis of the assessment objectives for GCSE Economics and Business qualifications. It is a type of assessment tends to be based on quantitative evaluation of student learning; it involves awarding marks for the number of points made, whether the student applies knowledge, whether there is balance in relation to arguments put forward and different perspectives presented and relates these to a number which when added together provides the assessor with the means to make a judgement about the learning of the student in relation to others. Those with a higher mark will be deemed to have demonstrated more of the learning objectives than those with lower marks.

In keeping with the revised assessment objectives for GCSE as established by QCDA, this regime is based on three assessment objectives:

AO1: Recall, select and communicate their knowledge and understanding of concepts, issues and terminology.

AO2: Apply skills, knowledge and understanding in a variety of contexts and in planning and carrying out investigations and tasks.

AO3: Analyse and evaluate evidence, make reasoned judgements and present appropriate conclusions.

For the two questions you are asked to mark, the assessment criteria will be given in the mark scheme to help guide your marking in terms of the weighting of the marks available related to each of the assessment objectives. Your assessment of the work, therefore, will require you to award marks according to the assessment objectives and then provide a total mark, out of 10 for Question One and 16 for Question Two.

Both questions form part of a range in an examination paper. The evidence/stimulus material available to the candidate in relation to these

questions has been reproduced for you so that you can see what information candidates had to help them prepare their answers.

For Question One, the evidence was presented to students at the beginning of the paper and before the set of questions.

In both cases, it must be borne in mind that students are answering these questions under exam conditions, they have limited time and limited space in which to give their answer. The emphasis, therefore, is not on how much the student writes but the way in which the student demonstrates the assessment objectives as outlined in the mark scheme. Remember, you are judging the responses against the assessment objectives targeted NOT what you could write as an answer, what has been missed out or what would constitute a 'perfect answer'.

Evidence:

Question 1:

Explain TWO reasons why a business such as Nokia might use the Boston Matrix.

(10 marks)

Knowledge (AO1) = 2 marks

Application (AO2) = 2 marks

Analysis (AO3) = 6 marks

1 mark for each reason (knowledge)

Possible reasons include:

- Monitoring its product portfolio
- Help Nokia plan ahead development of new products, for example

• To help Nokia make decisions - about launching or withdrawing a product

Consider possible investment plans

• Plan/respond in relation to competitors

1 mark for appropriate application to Nokia in each reason given.

3 marks for analysis of each reason. The analysis will make the link clear between the use of the Boston Matrix and the reason why Nokia might use it. Analysis will use appropriate terminology and may make reference to market share/market growth. Analysis may pick out the key features, causes, consequences, factors, reasons, as appropriate to the context.

Question 2: This particular question forms part of a series of sub-questions on the examination paper targeting different assessment objectives. This question was the last in this series of sub-questions and represents the high tariff question.

In this case, the evidence is presented in small sections with questions after each piece of evidence. Together these smaller pieces of evidence can be used by students to help support their answer in the high tariff question.

The Evidence:

The Olympic Bid

On Wednesday 6th July 2005, there were celebrations throughout the country as it was announced that the 2012 Olympic Games would be held in London. The estimated cost of putting on the Games is £2.4 billion (that's £2 400 000 000!)

Much of this money will be spent on building new facilities in Stratford, East London. The work involved in developing the sites for the Games will be carried out by organisations in both the private sector and the public sector

The cost of staging the Olympics will be high. Many cities who have staged the Games in the past have made very large losses and this has imposed a burden on taxpayers for many years after the Games have finished. The benefits of staging the Games therefore might not be as great as some people think.

Athens, which hosted the Games in 2004, suffered a loss estimated at £4.6 billion but Sydney who hosted the 2000 Games made a profit of £110 million

and generated £3 billion in tourism.

(Source: Adapted from Nick Goodway and Jane Padgham, Evening Standard, July 6th 2005).

The group of firms building the athletes village, will use the buildings after the Games have finished for retail and housing development. They estimate that the £3.6 billion investment spent on the site will bring a £5 billion return. The building companies had to bid against other building firms to win the right to build the athletes village.

(Source: Adapted from Nick Goodway and Jane Padgham, Evening Standard, July 6th 2005).

Not everyone is happy that London has won the right to stage the Olympics. Some people believe that the cost of hosting the Games is too high and that the money could be better used elsewhere.

The building of the new facilities including the athletes' village and the Olympic stadium will be in Stratford in East London. This is a relatively poor area of London. The investment will mean jobs for some, disruption for others and a boost for tourism for the whole of London and the UK.

Shareholders of businesses who win contracts to build new facilities for the Olympics might be hoping that the firms do indeed turn the investment into healthy profits which might boost the share price of the business but they will also be mindful of the delays and problems suffered by Multiplex, the building company responsible for building Wembley Stadium. Multiplex made a big loss on the project and the delays to opening Wembley meant its reputation also suffered.

The staging of the Olympics could provide thousands of jobs in many different areas. Hotels and other tourist attractions in and around London might benefit from the increase in tourism, transport users in the Capital may eventually benefit from a better transport system, businesses throughout the UK might benefit from the associated work that is likely to be created.

For example, the bid team suggests that clothing for 50 000 volunteers will Page | 323 have to be made - if such work goes to a UK textile firm it could be very beneficial indeed. Add into this the fact that in addition to hosting the Olympic Games, there will be the responsibility of hosting the Paralympics the benefits that will be given to those who have disabilities will be enormous as will the needs of disabled people who visit the city after the Games have long finished.

Source adapted from:

http://www.bized.ac.uk/current/mind/2004_5/221104.htm

The benefits to cities hosting the Olympics are mixed. Some argue that there are long term benefits for host cities. The last city to host the Olympics, Athens in Greece, saw the Greek government pay out approximately \$12 billion — more than twice the initial estimate — and the money spent by spectators' was not enough to make up the difference. Some Greeks did pretty well, especially in the construction industry, but the rest of the country was stuck with a steep bill.

Source: adapted from http://www.reason.com/hod/jw071105.shtml

Question 2:

Using all the evidence and your knowledge of business and economics, assess the strength of the case for hosting the Olympic Games in London in 2012.

(16 marks)

Knowledge (AO1) = 2 marks

Application (AO2) = 4 marks

Analysis and Evaluation (AO3) = 10 marks

You are required to mark this question using the levels of response mark scheme below:

Levels of response mark scheme:

Question Number 2. Indicative content

The question is designed to target analysis and evaluation skills. We are

Page | 324
looking for the student to present the case for and against London hosting the Games using appropriate terminology and employing concepts and methods expected of a student having studied economics and business at this level.

Having presented the case the student is expected to make a judgement in relation to the question asked.

Possible arguments in favour of hosting the Games include:

- Benefit to the whole economy
- Benefits to the local economy
- Creation of jobs

• Legacy of sporting venues and residential accommodation for use after the Games have finished

- Improvements in transport infrastructure
- Boost to tourist trade

Possible argument against hosting the Games include:

- Cost to taxpayer
- Potential for the country to end up in debt
- Opportunity cost of hosting the Games

• Impact on jobs and businesses in local area who have to close down or move

• The disruption and congestion caused by the construction programme

There is no 'right' answer here; the quality of the evaluation is the key to the answer. Analysis will pick out the key features, causes, consequences, factors, reasons, etc. as appropriate to the context.

Level 1 1 - 4

A judgement is made but the support for the judgement will be limited at the lower end of the mark band with a simple statement in support only. At the top of the mark band there will be some further development of the supporting statement given. Answers in this band will tend to be one-sided and make no use of appropriate terminology and concepts.

Level 2 5 - 8

A judgement will be made with some developed support offered which may be in the form of at least one source of analysis. At the lower end of the mark range the answer will be one sided but if students consider both sides of the argument the answer will tend to be in the upper level of the mark band. However, there will be limited analysis offered in this band which may be characterised by only identifying one source of analysis for one of the sides of the argument and nothing more than a judgement on the other. There will be no conclusion offered. There may be some use of terminology and concepts but these may appear confused or are used inappropriately.

Level 3 9 - 12

In this band a judgement is made with some developed support offered. Both sides of the argument will be presented and there will be some clear analysis offered for each although there may be more in favour of one-side than the other. A simple conclusion will be given which may, at the lower end of the level, be a re-statement of what has already been presented rather than an attempt to draw the analysis together. There will be some use of appropriate concepts and terminology which demonstrate some confidence or understanding of their value in this context.

Level 4 13 - 16

A judgement will be made and the level of support will be well developed. The analysis will be used to support the points being made and both sides of the argument will be given relatively equal weight. The student may use the 'it depends' rule when presenting their argument which will be supported by the placement of values on the points raised (it depends what level of debt is created or how many jobs are created etc.). The answer will be rounded off by a conclusion which relates to the question and draws on the analysis given. Answers in this level will demonstrate a command and facility in the use of terms and concepts.

Appendix XIII

SOLO Mark scheme explanation

SOLO Mark Scheme

This assessment regime is based on the SOLO Taxonomy. SOLO stands for Structure of Observed Learning Outcomes. The different bands which make up the taxonomy relate to increasingly complex structural responses. The first three bands describe quantitative change in learning outcomes whilst the latter two bands describe qualitative change.

Specifically, therefore, the taxonomy relates to two aspects of learning:

- 1. The learning of data some facts, skills, concept or problem solving strategies (quantitative)
- 2. The USE of these skills, facts or concepts in some way to solve problems, explain what has been learned, carrying out a task or making a judgement (qualitative)

The learning outcomes referred to in the acronym 'SOLO' describe what students will know and be able to do as a result of engaging in the learning process. For the qualification relating to the answers you are going to mark, these learning outcomes can be expressed as:

- Students will be able to recall, select and communicate knowledge and understanding of concepts, issues and terminology related to the topic area.
- Students will be able to apply skills, knowledge and understanding in a variety of contexts and in planning and carrying out investigations and tasks.
- Students will be able to analyse and evaluate evidence, make reasoned judgements and present appropriate conclusions.

When assessing the responses using SOLO, you are required to make a judgement about in which of the following five bands the answer is located. To use SOLO you need to view the answer as a whole rather than looking for Page | 328 particular points against which to award 'marks'.

In order to do this, you will need to read the answer through first to get a feel for the quality of the response. Next refer to the bands and look at the key factors that characterise each band. Then re-read the answer and note where the answer relates to the bullet points in the bands below. If your judgement of the answer reflects one or more of the bullet points in the bands below then it is likely that the answer will be located in that band.

At the end of the answer, write a **brief** explanation for your justification in relation to the band chosen.

Band 1 answer.

- Answer given is a tautology the question is simply repeated.
- A judgement may be given but with no supporting argument and simply restates the proposition in the question.
- No evidence of any understanding of the question the point may be completely missed
- The answer will have no structure and may not contain any evidence.
- Some knowledge may be recalled but it is by rote and no connections will be made between the information given
- Information given may make little sense and have no organisation
- Reference to evidence may be made but nothing is added or the focus may be on an irrelevant piece of evidence.
- If a conclusion is given it is based on transduction the conversion of information to another form without adding to its meaning or understanding.
- This transduction may also be an anecdote from the student's experience but is irrelevant to the question or the context.
- Any support given to a conclusion will be irrelevant.

Band 2 answer.

- Some minimal understanding of the topic/issue/context may be evident
- The answer may be a simple unqualified response on one aspect of the data/evidence/context/issue
- The answer may focus on terminology with no development
- Terminology may be used but is inappropriate or confused
- The answer may focus on just one piece of the evidence.
- Only one part of a task will be tackled or the answer will just be single focus
- Simple and obvious connections may not be made
- If any connections are made, the significance or relevance of them is not grasped
- A judgement may be given where appropriate but only one piece of evidence is used to support it.
- Judgements may be given which are repeated from the evidence and is likely to be the dominating or very obvious view in the evidence.
- A conclusion may be dogmatic or simply repeat the point already made.

Band 3 answer.

- Two or more elements of the questions will be addressed
- There may be a number of concepts or issues addressed but these will be disorganised.
- Factors/issues addressed will be treated as a sequence of isolated events.
- Some understanding of issues are evident but these will remain discrete - the trees are seen but not the wood!

- Whilst more than one issue etc. is addressed any conflicts or inconsistencies between them are ignored.
- Some use of the evidence to help demonstrate some understanding but the evidence is not used to infer or develop an argument.
- Examples may be used to highlight points made but only serve to have the characteristic of having multiple goes at the concept/issue with no links made.
- Pros and cons/advantages and disadvantages/costs and benefits are not weighted or assessed to give a balanced or informed conclusion.
- The possibility of an alternative explanation/perspective may not be recognised and so the conclusion is single focus.
- Any conclusion given replicates the predominant view in the evidence/context but does utilise some facts from the evidence.

Band 4 answer.

- Significance of the parts in relation to the whole recognised the trees and the wood are seen!
- Understanding of several issues/elements/components which are conceptually integrated.
- There is a relationship recognised between facts, theory, action and purpose.
- Facts/points are used to address the question and help make sense of the topic as a whole.
- Overall structure of the answer is coherent.
- Some points will have balance which demonstrate a recognition of different perspectives related to the context/evidence/theory/concepts.
- The main issue will be taken and a contrast/comparison made with other related issues.

- The main concept will be recognised (i.e. opportunity cost, market failure, productivity, efficiency etc.)
- Both sides of the argument are presented and related to appropriate concepts.
- Pros and cons/advantages and disadvantages/costs and benefits will be weighed up in arriving at a conclusion.
- Analysis given in a coherent manner consisting of more than two linked factors/issues/causes/consequences/issues.
- 'It depends rule' unlikely to be used in assessing these factors etc.
- Conclusion will be given which is partly considered but does not go outside the context.
- Conclusion/judgement may be given at the outset and then an argument presented in support which may lead to an unbalanced conclusion.

Band 5 answer.

- Answer is rooted in a general principle/concept rather than a given context/concept.
- Knowledge/concepts/theories are applied and generalised to new and unfamiliar contexts and situations.
- Knowledge/concepts are used seamlessly in the argument there is a focus on using the knowledge rather than trying to demonstrate volume of knowledge.
- Connections with the subject area made but also outside it.
- Conceptualisation takes place at a level beyond that which would be expected from normal teaching of the topic/issue and may include abstraction.

- The analysis given is also used to make generalisations of behaviour and theory showing a transfer of principles and ideas underlying the topic.
- Links are made between conflicting pieces of evidence. Logical inferences are drawn from the context/evidence/analysis.
- The answer may show some evidence of empathy with players in the evidence/context not drawn simply from the student's own experience.
- Perspectives and analysis are given in a relative context using the 'it depends rule'.
- Answer may start with a hypothesis, then a consideration of the evidence. The question is held open whilst the evidence is considered.
- A reasoned argument is presented which combine the variables identified and analysed to form a comprehensive judgement.
- The student distances themselves from the judgement to provide an objective conclusion/judgement which may be arrived at through deductive reasoning.
- Information /contexts/concepts are brought in from outside the immediate topic to inform judgements and/or conclusion.
- Judgements/conclusion may be relevant to the question but the student also sees that a different conclusion could easily be reached with some minor adjustments to perspective/analysis.