The difficulties of conceptualising and operationalising critical thinking through assessment practices: a case study of academics in a UK business school

D'NORTHWOOD, GAVIN

The difficulties of conceptualising and operationalising critical thinking through assessment practices: a case study of academics in a UK business school

Gavin D'Northwood

Abstract

Critical thinking defines a university education. If universities are nurturing critical thinkers then this should be apparent from and demonstrated through assessment. This is challenging, however, because critical thinking is subject to wide interpretation, with little or no academic consensus as to its specification. Additionally, assessment in higher education has long been a challenging issue, particularly so for complex higher-order thinking. Research continues to demonstrate considerable variability across assessors and the literature affords many reasons for this, but hitherto unexplored is the potential role played by differing conceptualisations of critical thinking as applied to the task of assessing students' work. In response, this thesis examines how participants - academics in a UK university business school - conceive of critical thinking in application to assessing undergraduate dissertations.

A case study methodology is employed, with a qualitative approach to data collection (via semi-structured interviews involving document elicitation) and thematic analysis, underpinned by social constructionism. The findings illustrate the perceived facets of an undergraduate student as critical thinker, and identify how individualised approaches to assessment differ, questioning the effectiveness of mitigating institutional mechanisms. This thesis adds to our theoretical understanding of critical thinking in higher education, in demonstrating that participants conceptualise critical thinking as a set of skills, dispositions and originality, and in proposing a link between models of epistemological development and academics' assessment related expectations of students' critical thinking. Further practical contributions are offered through demonstrating what participants look for as evidence of students' critical thinking and how this can influence assessment, together with the challenges to achieving consistency and averting grade variation. These could be of use to institutions, academics and students, subject to considerations of transferability.

The difficulties of conceptualising and operationalising critical thinking through assessment practices: a case study of academics in a UK business school

Gavin D'Northwood

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Education

School of Education

Durham University

2021

Table of contents

List of tables

List of Figures

Declaration

The material contained in this thesis is solely attributable to the author. No part of this thesis has previously been submitted for a degree in Durham University or any other institution.

Statement of copyright

Acknowledgements

I offer immense thanks to my supervisor team of Professor Julie Rattray and Dr Jonathan Tummons for their unceasing support and learned advice throughout the currency of this thesis. Special mention too for their forbearance over the good times and the bad, as I have wrestled with the complexities of managing part-time doctoral research with full-time employment and a full family life in a tumultuous world.

For all the additional help and support given by friends and colleagues, and for my participants for agreeing to do just that, I offer my gratitude.

Thanks also to my wife, Nicola, and our two boys who have grown up with this research and have stuck with me regardless. I am most grateful for their continued support and understanding.

1. Introduction

In this thesis I report on my doctoral research project in which I have investigated how academics in a UK university business school conceive of 'critical thinking' in application to their assessment of undergraduate dissertations. There is a personal inspiration for this project. It comes from my own transition to academia from professional accountancy practice. I recall being immediately 'thrown in the deep end' of teaching students and assessing their work with little or no training or guidance to fall back on. I was therefore heavily reliant upon advice from more experienced colleagues. In particular, I remember receiving my first tranche of assignments to assess. Having worked through some of these with colleagues for the purposes of calibration I found myself picking up and repeating on their assessment language. The more assignments I assessed, the more I came to realise that in my written feedback I was repeating particular phrases *ad nauseam* which I had apparently acquired from said colleagues. Such phrases included 'lacks criticality', 'could demonstrate greater critical appreciation' and 'needs more critical thinking'. I had read these phrases in colleagues' feedback to students, had heard them uttered in discussion and had seemingly absorbed them. I began to wonder what these apparently interchangeable stock returns actually meant and if, as academics, we were all effectively on the same page with our understandings. I did consider, as I do now with more confidence, that I knew what I was referring to and why. I also felt that I was not alone in this. Yet this left me wondering whether such understandings are shared across academics and, in particular, if we are all applying the same concept of 'critical' when assessing students' work.

What follows is a more detailed rationale for this study. In this, I provide relevant background formed with reference to views coming from the literature on critical thinking and assessment in higher education. I then frame the parameters of this study through consideration of research objectives, questions and the case setting in which I situated this research. I then acknowledge subjectivity and consider of my own reflexive position in response. Finally, I provide a chapter-by-chapter outline of how this thesis is organised.

1.1 Background to the study

1.1.1 Critical thinking in higher education

In 1990, the American Philosophical Association declared critical thinking a vital underpinning to our democratic society (Lorencová, Jarošová, Avgitidou & Dimitriadou, 2019). Three decades on our current world context is now even more complex, characterised as it is by the global knowledge economy, rapidly changing technology, political turmoil, economic crises, social volatility and climate change (Penkauskiene, Railiene & Cruz, 2019; Roohr, Olivera-Aguilar, Ling & Rikoon, 2019; Schendel & Tolmie, 2017). In 2020 and 2021, we have also witnessed first-hand the complex political, social and economic difficulties caused by a global pandemic. We live then in a world that is both complex and subject to continual change, a world beset by difficult problems in need of complex solutions (Kek & Huijser, 2011). It is in this context that critical thinking is rightly labelled "an essential skill for 21st century survival" (Luk & Lin, 2015, p.67). This I submit foregrounds the importance of critical thinking in education.

Critical thinking is "arguably the very essence of what higher education is meant to inculcate in students" (Rear, 2019, p.665). Indeed critical thinking has become the very hallmark of higher education and traditional university degrees, a concept that defines higher education no less (Barnett, 1997; Lincoln & Kearney, 2019). Critical thinking is thus widely promoted as a key educational aim (Bok, 2006; Luk & Lin, 2015; Pithers & Soden, 2000). Universities view developing students' critical thinking as a primary objective; often professing that their students when they graduate will do so as critically reflective beings (Golding, 2019). Business schools are particularly concerned with developing critical thinkers, described as graduates who can "manage complexity, uncertainty, equivocality and value conflicts" (Rodriguez, 2009, p.523). The emphasis on critical thinking in higher education is evidenced by the prominence it is given in teaching, assessment and feedback. Critical thinking is embedded in the language of educational standards at all levels, and prioritised as a primary educational policy goal through the emphasis placed on graduate attributes and employability skills (inclusive of critical thinking) (Barrie & Prosser, 2004; Carrington, Chen, Davies, Kaur & Neville, 2011; Cosgrove, 2011; Espey, 2018; Leong, 2013; van der Zanden, Denessen, Cillessen & Meijer, 2018). Subject knowledge taught at university will inevitably wane or

be superseded but the ability to think critically will endure (Terenzini, Springer, Pascarella, & Nora, 1995).

In UK higher education, critical thinking is a key stated graduate outcome (Brennan & Osborne, 2005). Its significance to UK policy makers is made clear by the Department for Business, Innovation and Skills (BIS) (2016, p.5): “The skills that great higher education provides – the ability to think critically and to assess and present evidence – last a lifetime and will be increasingly in demand.” By virtue of granting degree awards, UK higher education institutions are necessarily confirming that students have met the learning outcomes for their programmes of study, learning outcomes that invariably flow from the Quality Assurance Agency (QAA) via the UK Quality Code for Higher Education – The Frameworks for Higher Education Qualifications of UK Degree Awarding Bodies October 2014. Here, there is clear support for critical thinking as a required outcome. For example, for level 5 qualifications (foundation degree) students must demonstrate “knowledge and *critical* understanding of the well-established principles of their area(s) of study”; the “ability to evaluate *critically* the appropriateness of different approaches to solving problems in the field of study”; and the ability to “undertake *critical* analysis of information” (p.23). Level 6 (honours degree) adds an expectation that qualification holders will possess the necessary skills of evaluation vis-à-vis “arguments, assumptions, abstract concepts and data…” (p.26). Thus, the importance of critical thinking is espoused, flowing from the QAA to HE institutions, evidenced also by the prominence of the word ‘critical’ and associated terms in the mission statements, programme literature, qualifications and grade descriptors of institutions (Moon, 2008).

Support for critical thinking is not confined to policy makers. Academics also value critical thinking in their students (Leong, 2013; Moore, 2013). For example, over 90% of faculty members in one US-based survey declared the key goal of undergraduate education to be the development of students’ critical thinking (see Bok, 2006). Similarly, in a UK-based study in 2013, participating academics were clear in articulating a belief that critical thinking is central to their practice as educators and that students need to evidence this, albeit without those participating academics necessarily demonstrating a shared understanding of what critical thinking actually involves (see Moore, 2013).

Critical thinking as a graduate attribute is also highly prioritised on employability agendas, important by the estimations of employers and students (Hinchcliffe & Jolly, 2011; James, Hughes & Cappa, 2010; Leong, 2013; Tuononen, Parpala & Lindblom-Ylanne, 2019). However, talk of critical thinking in this context narrowly considers it as a skillset. Featuring prominently here are skills of decision-making and practical problem solving (see, for example, Macpherson & Owen, 2010; Wilson & Howitt, 2018). These are apparently what employers' prize most highly in graduates. Yet I will demonstrate through Chapter 2, the first of a two-part literature review, that critical thinking is a multi-faceted concept that amounts to much more than this, encompassing skills, dispositions and more (Moon, 2008; Davies, 2015). Barnett (1997) in particular would find the equating of critical thinking with employability skills as unnecessarily limiting, arguing that the purview of higher education should be the development of people who are critical beings, i.e. people who, being self-aware and reflexive, can through critical reasoning make critical decisions and take critical actions that benefit society. To focus on problem solving, which Barnett would regard as occupying the lower levels of critical thinking, arguably limits the development of truly critical thinkers. Nonetheless, observably much of the literature on critical thinking in higher education is focused on critical thinking as skills.

Evidently, in higher education we expect students to move beyond passive reception of that which is given as known, pushing through to active engagement in higher-order thinking, even creation of new knowledge (Bok, 2006). This should be apparent from and demonstrated through evaluation of learning outcomes (Arum & Roksa, 2008; Klein, Benjamin, Shavelson & Bolus, 2007; Lee, Lee, Makara, Fishman & Hong, 2015). It is axiomatic that students who achieve higher grades necessarily demonstrate greater critical thinking (Dwyer, Hogan & Stewart, 2014; James, McInnis & Devlin, 2002; Stassen, Herrington, & Henderson, 2011). Yet the substance of critical thinking as concept is less than explicit in academic settings (Wendt & Ase, 2015).

Critical thinking then is a difficult term to get to grips with (Williams, 1976). Although used pervasively in higher education, particularly with respect to assessment, it is not precisely specified nor clearly conceptualised (Baril, Cunningham, Fordham, Gardner & Wolcott, 1998; Harrell, 2011; Kuhn, 1999; Leong, 2013; Moore, 2013; Wendt & Ase, 2015). Subject to wide interpretation, critical thinking has been defined in various terms carrying numerous meanings following from Glaser's 1941 seminal piece *An Experiment*

in the Development of Critical Thinking. Critical thinking is intuitively learned by academics who claim to recognise it when they see it but who cannot easily define or explain what it is (Atkinson, 1997; Fox, 1994). There is no universally defined list of what constitutes critical thinking and no academic consensus as to its specification (Harrell, 2011; Moore, 2013). It is not surprising then that critical thinking is apparently not well understood by academics (Moore, 2013).

1.1.2 Assessment in higher education

High stakes assessment is a central function of higher education and a crucial factor in determining future study and career opportunities (O'Hagan & Wigglesworth, 2015). This is in the context of the modern landscape of higher education, which is characterised by greater global student mobility, international competition for students, rising student numbers, larger class sizes and increasing pressure on student/staff ratios (Gu & Schweisfurth, 2011; O'Connell et al., 2016). Accountability and transparency are especially important in this context (Wendt & Ase, 2015). As a result, higher education assessment practices are located within an agenda of institutional accountability underpinned by quality assurance frameworks (Bloxham, Boyd & Orr, 2011). These cover the design and execution of assessment tasks and associated processes for evaluating and moderating student performance leading to justifiable assessment decisions backed by evidence (Bloxham et al., 2011; Grainger, Adie & Weir, 2016; Grainger, Purnell & Zipf, 2008).

Producing any good piece of lengthy written work involves a student demonstrating skills of selection and evaluation in forming cogent arguments backed by supportive evidence, written within the customary academic genre, and suitably attentive to relevant academic conventions. This is a complex achievement that resists measurement and is therefore best assessed through judgement (Yorke, 2011). However, key concerns within the agenda of accountability are to reduce the apparent arbitrariness of assessors' judgements and ensure that grades fairly represent student achievement (Sadler, 2009; Yorke, 2010). Professional judgements along the lines of 'I know a 2:1 when I see it' (Ecclestone, 2001 refers) are hard to justify in this context. Moreover, researchers continue to point to deficiencies in assessment systems in education settings; deficiencies which render universities seemingly incapable of producing comparable

grades between assessors on a reliable and consistent basis (Bloxham, den-Outer, Hudson & Price, 2016a; O'Connell et al., 2016).

Significant efforts have been made to improve fairness in assessment through greater transparency (Jonsson, 2014). Publication and consistent application of appropriate assessment criteria are pivotal here (Bloxham et al., 2016a). Assessment criteria and rubrics or grading schemes are said to help assessors and students by clarifying standards of achievement at different levels, promulgating shared understandings, facilitating transparency and hence obviating assessors' prejudices (Menendez-Varela & Gregori-Giralt, 2016). However, irrespective of how well defined criteria may be, research has demonstrated considerable variability in assessment, i.e. assessors applying the same criteria may still arrive at different grades for the same piece of work (Bloxham, 2009; Grainger et al., 2008). The literature offers a variety of reasons for this, which I cover in more detail in Chapter 3. These, however, do not specifically question the potential role played by differing conceptualisations of critical thinking held by academics as applied to the task of assessing students' work.

I reason that a varied academic base means conceptions of what is required or valued in dissertations is likely to vary, more specifically what critical thinking is and how it manifests in dissertations is likely to vary, and that this could affect assessment. When assessing students' work, academics must hold the right concept of quality (of which critical thinking forms a part) for the assessment task and judge the work in relation to that concept (Grainger et al., 2008; Popham, 2005; Sadler, 1989). Academics must appreciate what critical thinking is in order to judge it. We know that critical thinking is difficult to define. It follows that it is difficult also to assess. Whilst assessing students' content knowledge is relatively straightforward, assessing critical thinking most certainly is not (Bissell & Lemons, 2006). Is critical thinking, as Ebel (1965) suggested, simply too intangible an outcome to measure? Is critical thinking a concept that simply "cannot be assessed" (Bissell & Lemons, 2006, p.66)? Alternatively, is critical thinking, as Ennis (1993) suggested, something that is difficult to assess well but is nonetheless assessable?

If it is possible to assess critical thinking then, agreeing with Golding (2019) in relation to thinking more broadly, the fundamental assessment issue here seems to me to be one of discernment. How do academics actually discern when, how and to what extent critical thinking has occurred? How do they know what to look for? How do they recognise it? How do they evaluate it? How do judge its worth? How does this then affect the grade attributed as the assessment outcome?

1.2 Research parameters

1.2.1 Research objectives and questions

From the foregoing, it would seem that major questions of what critical thinking actually is for academics and how they perceive and judge it when assessing students' work remain unresolved. It is a given that the notion of critical thinking is important in higher education but what exactly is it, how is it understood, how is it discerned and how is it assessed? The uncertainty here points to a significant gap in the literature. There is a clear need for determining how academics understand what critical thinking is and how they operationalise this through assessment. These form the overall objectives for this study. Pursuant to meeting these, I constructed three broadly constituted research questions, RQ1-RQ3 inclusive:

RQ1: How do business school academics conceptualise critical thinking?

RQ2: How are business school academics' conceptualisations of critical thinking operationalised through assessment practices?

RQ3: What are the challenges to achieving consistency when assessing for critical thinking in undergraduate students' dissertations?

1.2.2 The case study

This research is located within a university academic faculty, specifically a business school. The university in question, a member of the Russell Group of British research universities, is a well-established, traditional and research-intensive city-based university. The host business school is highly ranked nationally and internationally for the quality of its research and for the standard of its academic programmes. It is one of an elite group of UK business schools holding triple accreditation from three internationally recognised quality assurance bodies: the Association of MBAs (AMBA), the European Quality Improvement System (EQUIS) and The Association to Advance Collegiate Schools of Business (AACSB). The school offers courses on undergraduate and postgraduate degree programmes in the disciplines of accounting, economics, finance, business and marketing. It has in the region of four thousand students in any given year and employs upwards of a hundred and seventy faculty members (FTE) across the subject disciplines (excluding PhD students) (2020 figures).

As the host business school is part of a research-focussed institution, the dissertation module, through which students demonstrate that they can undertake a significant piece of independent research, then takes on added importance. Dissertations provide the capstone of an undergraduate programme of study (Ashwin, Abbas & McLean, 2017; Webster, Pepper & Jenkins, 2000). This type of capstone module is a well-established feature of undergraduate degrees in the UK. Positioned at the culmination of a degree programme, it provides students with an opportunity to synthesise what they have learned and demonstrate their independence and higher-level capabilities including their critical thinking (Hammer et al., 2018; Lee & Loton, 2019). That dissertations are open-ended and lack certain outcomes serves to encourage critical thinking in students (McQuade, Kometa, Brown, Bevitt & Hall, 2020). The student chooses the topic of investigation, thereby defining the focus of the piece, and conducts the investigation under academic supervision. The investigation necessitates collection and analysis of data, whether primary and/or secondary based, over an extended period. Students can show their maturity as independent learners through completing the most complex, substantial and independent piece of work required of them from their degree programme, and in so doing exhibit their readiness to graduate (Berheide, 2007; Bettany-Saltikov, Kilinc & Stow, 2009; Lee & Loton, 2019; Todd, Bannister & Clegg, 2004).

In this thesis, I employ undergraduate dissertations as a vehicle for examining how academics understand and operationalise critical thinking in assessment. All undergraduate programmes at the host business school require students to complete a research dissertation of up to twelve thousand words as a double-weighted final year capstone module. The module, including its constituent requirements and assessment guidance, is common across all programmes. The dissertation is a substantial piece of work involving detailed and critical examination of a particular topic chosen by the student. Empirical research is encouraged but not compulsory. Each student is allocated a supervisor who provides guidance and support through to submission via six supervisory meetings structured periodically throughout the academic year. The supervisor also assesses and offers feedback on a draft dissertation chapter as a piece of formative work.

Performance is assessed against criteria. These, together with relevant policies and procedures, are published to staff and students in a dissertation module handbook. The host business school employs a generic set of grade descriptors applicable to all modules on undergraduate programmes in lieu of more detailed 'assessment criteria' as interpreted in the literature (I expand upon this in Chapter 2). These grade descriptors are the only criteria given to assessors to apply to the task of assessing undergraduate students' work, including final-year dissertations. I understand this type of approach is common within pre-1992 universities. The grade descriptors (Appendix A) are split into bandings: First Class (70+), Upper Second Class (60-69), Lower Second Class (50-59), Third Class (40-49) and Fail (<40). First Class is segmented by Exemplary (86-100), Outstanding (76-85) and Excellent (70-75); Upper Second Class by Very Good (65-59) and Good (60-64); Lower Second Class by Adequate (55-59) and Fair (50-54); Third Class by Weak (45-49) and Poor (40-44); with Fail grades split across several segments ranging from Very Poor to Outright Fail.

The word 'critical' first appears in Good (60-64) described thus: "Reasonably good knowledge and understanding, but little evidence of critical assessment or analysis. Coherent presentation but less well-structured than seen at higher grades." Very Good (65-59) sees the removal of "reasonably" so that the requirement is for "good knowledge", prefaces "understanding" with "thorough", and adds a requirement for "evidence of broader understanding informed by wider reading". Answers categorised as

Upper Second Class are demarked from First Class through students displaying a “less critical grasp of the subject”. As we move up the scale the requirement for the level of “knowledge” moves from “good” (65-69) to “detailed” (70-75), “comprehensive” (76-85) and “complete” (86-100). “Evidence of judgement in selection and critical analysis of relevant material” features at both 70-75 and 76-85. “Minor errors” are “acceptable” “if compensated by excellence in other areas” (70-75). Errors are not mentioned above this, implying that answers at or above 76 will be error-free. Outwith purely quantitative research this aspect must be incredibly difficult to achieve or indeed to assess. What, for example, is error-free when it comes to a subjective piece of interpretative research? “Structure” features at the First Class level, moving from “good” (70-75) to “logical” (76-85). “Depth of understanding” further demarks 76-85 from 70-75. Work of Exemplary standard (86-100) shows “insight”, which is not referred to below this level.

The assessment criteria are there to aid consistency. However, assessment criteria cannot achieve the necessary shared understandings that would underpin consistent application simply by virtue of their existence (Menendez-Varela & Gregori-Giralt, 2016). Communication, exchange and negotiation are pivotal. Hence, all undergraduate dissertation supervisors/assessors receive an invitation to attend an annual meeting at which the dissertation module leader details the policy and requirements for the module, including its assessment criteria. Having witnessed such a meeting I interpret this to be a didactic session, with little or no space given over to discussion. The assumption is that once communicated, anyone can apply the documented assessment criteria with the requisite consistency. Anecdotally, attendance at this annual meeting by supervisors is patchy at best. I would back this observationally, albeit from only one such meeting that I attended. I did find the meeting to be informative but limited as a vehicle for facilitating the sharing of knowledge and understanding.

As a further aid to consistency, all undergraduate dissertations at the host business school are internally moderated via a process of ‘double marking’, also known as ‘double appraisal’ (per Sadler, 2013), whereby two academics assess a dissertation separately and without reference to each other. The two then meet to discuss their respective grades with the aim of reaching a consensus grade representing the student’s overall achievement vis-à-vis the criteria. That the two assessors meet to compare and discuss grades is arguably beneficial in that pairs of assessors can work through and jointly agree on the application of the assessment criteria to a dissertation leading to a finalised grade. This affords assessors opportunities to discuss and develop their knowledge of

the criteria as applied (Beutel, Adie & Lloyd 2017; Bird & Yucel, 2013; Price, 2005). This is better than and acts as a check on one assessor applying criteria, whilst it may further help to make individual tacit knowledge more explicit. The aim of such a practice is to build fairness, consistency, justification and defensibility into grading decisions (Bloxham, Hughes & Adie, 2016b). This, as with comparable internal moderation mechanisms, then offers inbuilt safeguards for institutions, assessors and students (Bloxham & Boyd, 2012; Hand & Clewes, 2000; Partington, 1994; Webster et al., 2000).

However, in such a large business school separated along disciplinary lines there are inevitably a great many pairs of assessors leading potentially to a great many differences in interpretation and application of criteria. This means that the grade awarded to a dissertation is still materially subject to who actually assesses it. The assessment policy includes the requirement that should first and second grades differ by ten or more percentage points, or split a grade boundary, then the first assessor must complete a form explaining the difference and any action taken to come to a consensus view. Absent consensus, recourse to a third arbitrating assessor is made. Samples of dissertations from all grade band levels are referred to external examiners as a further check on the consistency and fairness of the assessment process. That these aspects of the policy exist at all recognises that differences do occur and need to be controlled for. Analogous to this, research has shown substantial yet seemingly inexplicable variation and limited reliability between first and second assessors (See, for example, Brooks, 2012; Cannings, Hawthorne, Hood & Houston, 2005). Moreover, whilst Kuzich, Groves, O'Hare and Pelliccione (2010) would maintain that double marking processes serve to improve accuracy and inter-rater reliability, Bloxham (2009) and Cannings et al. (2005) would appear to contend otherwise. Whether two assessors can come together and agree on a grade that is better and more accurate than the sum of its parts is highly debateable therefore.

1.3 Subjectivity

This is my thesis. I am the researcher. I am also a business school academic. I do not imagine that I can escape subjectivity here. This is inevitable given my immersion in the social world under investigation and through the way in which I conceived and executed this research. I therefore do not seek objectivity or claim to have achieved it. Instead, I have chosen to embrace and account for subjectivity as best I can, recognising that I am part of and therefore inextricably linked with the social world I am studying. Subjectivity is unavoidable, hence the imperative for demonstrable rigour throughout the research process and for vigilance through reflexivity on my part (Farquhar, 2012).

1.3.1 Reflexive statement

In embracing subjectivity, it is important that I am able to explore reflexively my relationship with the research (Brannick & Coghlan, 2007). I have employed a case study methodology, within which I have taken a qualitative approach to data collection (via semi-structured interviews) and thematic analysis, underpinned by social constructionist research philosophy, to examine a contemporary social practice situated within a specific setting. From a social constructionist viewpoint all meaningful reality is socially constructed (Saunders, Lewis & Thornhill, 2012). People as the 'objects' of this social world, unlike the objects of the natural world, are able to attribute meaning to their environment (Bryman, 2015). We actively construct the meanings that we attribute to our world as we engage with it (Crotty, 1998). Given the constructed nature of reality there ought to be differences, contradictions, struggles as to meanings and hence contested versions of reality which I as the researcher must seek to bring to the surface (Cohen, Duberley & Mallon, 2004). My aim then is to develop an understanding of my participants' subjective realities, to access the world through their eyes, forming empathic understandings (Bryman, 2015). Appreciably, however, the picture I develop will inevitably be subject to my own perspectives and biases as a social actor. I must be critical then of my own place in this, aware of imprinting my own frames of meaning on the reality that I construct from the research data (Cohen et al., 2004). The case study methodology can help me with this, challenging rather than simply confirming my own biases, noting that the greater bias for Flyvberg (2011) is toward falsification rather than verification of preconceived notions.

That I as the researcher am encouraged to be reflexive is a key strength of social constructionism (Johnson & Duberley, 2000). This translates as the need for me to be explicit about my own personal values and perspectives that could influence the study and to explore how my biography may have shaped it (Burr, 2003). I recognise that I am "integrally involved in the case" (Cohen et al., 2017, p.376). I am central to the way in which I have conceived and executed this study and necessarily complicit in the construction of the version of social reality presented (Cohen et al., 2004). I cannot pretend to be a detached observer, or that I have had no influence on the research, but what I can do is be reflexive, acknowledging my presence in the process (Hackley, 1998).

I should then declare my prior experience of the practice under investigation (Malterud, 2001). I hold certain assumptions about the processes, practices and beliefs inherent to assessing dissertations, the importance of assessing for critical thinking in this and of how critical thinking is conceived and judged. This comes from my own experience. I began this chapter with an exposition of how the idea for this study formed from my transition to academia from professional accountancy practice. Prior to joining academia I had enjoyed a career spanning sixteen years in administration, auditing and financial management. I have been an academic since 2007, in which time I have taught and assessed students at undergraduate and postgraduate levels on a variety of accounting, finance, business and law related modules at two universities. I have supervised and assessed hundreds of research dissertations, predominantly on undergraduate degrees but also postgraduate degrees. I have also worked as an examiner for two professional accounting bodies and as an external examiner for universities other than my employer. I have then extensive and relevant experience of assessment practices.

I conducted this research as a full-time academic and a part-time doctoral student. This insider dual-status brings some advantages but also presents certain issues. I have the advantages of pre-knowledge, experience and understanding of the setting (Bell, 2014; Saunders et al., 2012). I have the further advantage of pre-existing relationships with participants, albeit to varying degrees of contact and closeness, affording both approachability and goodwill. These proved useful in gaining primary access to the setting for research purposes and in building a sample of willing participants. Appreciably these came largely from what I would describe as the circle of the willing, in my estimation a motivated subset of the academic faculty of the host business school. Regrettably, some requests for engagement from outside this circle were as anticipated

either rebuffed or ignored. For richness, it would have been of value to include other voices, particularly those few anecdotally difficult characters who could have added different perspectives to the mix. As for those who did choose to participate, I know several of those involved, can estimate what they are thinking and can draw on my own experience in interview discussions to tease out responses, all the while accepted as an insider and hence, above suspicion, a colleague, a confidant and in some cases a friend. On the downside, I am of course open to accusations of being too close to the setting, of making assumptions and not probing deeply enough, of being incapable of observing the objectivity needed for valid research. However, this view is challenged by Brannick and Coghlan (2007) who argue that insider research can be valid and can yield useful information on organisational reality which traditional research approaches may struggle to unearth. Whether I am not sufficiently distant from the setting and hence, incapable of being objective is in any event moot. I have stated that I have embraced subjectivity and I have further set out my position on the rigour of this research in Chapter 4 (s.4.8). I accept that there are also ethical issues to contend with, not just in terms of maintaining confidentiality, etc. but also in recognising that interviews could get uncomfortable at times and that this may affect my relationships with participants going forward (Bell, 2014). Thankfully, whilst interview discussions ranged and I believe many honest things were said, I genuinely cannot recall any particular instances of discomfort and I have not observed any adverse effects on relationships.

I have taken care to critically reflect upon my own stance, the “framework of meanings” which I hold (Hamilton & Corbett-Whittier, 2012, p.53), borne of my assumptions, my own constructed version of reality coming from my own experiences, etc., so as to avoid being too influenced when interpreting the data (Cohen et al., 2004; Vaismoradi et al., 2016). I recognise that my background shapes my stance. Brought up in a staunchly religious family I am a practising Catholic. I have a degree in law. I am professionally qualified as an accountant. I have worked predominantly in the public sector. I hold to the importance of integrity, honesty, morality, justice and fairness. These manifest in the context of assessing undergraduate dissertations as a desire to assess as objectively and fairly as is possible. From experience, I would hold that others have varying perspectives and approaches to assessing undergraduate dissertations dependent upon their own backgrounds, sets of beliefs and priorities. Some of whom I have double marked with, and so brought assumptions as to their respective positions and approaches. Others I have not, and so I have no pre-conceived notions except perhaps general ones based on whether the participant was more or less research engaged,

epistemologies, methodological preferences, etc. Some participants' views chime with my own, some jar, which could potentially cause disagreements and conflict.

Embracing subjectivity notwithstanding, I looked to bracket out any pre-conceived notions I may have had about participants and their respective approaches to assessing undergraduate dissertations. Bracketing requires me to set aside any relevant views and preconceptions before tackling the data premised on the concern that these may influence and shape the data analysis (Tufford & Newman, 2012). I accept that true bracketing is virtually impossible. What I tried to do then before analysing the data is to access and reflect upon my tacit knowledge of the organisation and of participants, deeply ingrained through socialisation in the setting (Brannick & Coghlan, 2007), supported also by thematic analysis processes of decontextualizing and amalgamating utterances on like areas, which helped to take the individual out of it.

I am acutely aware of my position as the primary research instrument (Mathews & Ross, 2010). I, as the researcher, produced the data for this thesis through interactions with participants (Cooper & Morgan, 2008). I recognise that qualitative interviews are co-constructed by researcher and participant (Brinkman & Kvale, 2015). It follows that participants' responses will be affected by this interaction. Hence, different researchers would generate different interactions co-constructing different accounts. Experience suggests that research interviewing is demanding (Coolican, 2009). It is a craft, a complex set of skills to be learned and honed through practice and experience (Kvale, 2007; Rubin & Rubin, 2012). I possess substantial experience of interviewing from an auditing background but not experience of research interviewing *per se*. Pilot interviews it is said can aid in developing interview techniques, particularly for those who are new to research (Sin, 2010). I would agree that the two pilot interviews I conducted did help with this. Moreover, I benefited from being an experienced business school academic, possessing then extensive familiarity with the "theme and context of (the) inquiry" which Kvale (2007, p.49) would regard as preconditions for "expert interviewing". I found the semi-structured format helped me to keep interview discussions on-topic, whilst giving participants the room to explore nuances and personal experiences connected with the practice under study.

I chose how to code and ultimately thematise the data, decontextualising and recontextualising it in the process (Nowell et al., 2017). I sought to unearth the experiences and understandings of participants. I wanted to be “a faithful witness” to their accounts (Novell et al., 2017, p.5). However, appreciably these emerged filtered through the lens of my own stance, formed from my personal perspectives, pre-existing thoughts and beliefs (Starks & Trinidad, 2007). Following the example of Braun and Clarke (2012), my thematic analysis combines inductive and deductive approaches. Inductive because I have coded directly from the data ‘bottom-up’, seeking therefore to access participants’ experiences, and deductive because I also draw on aspects from the literature and apply these ‘top-down’ to help me interpret and make sense of those experiences. In essence, I necessarily approached the data with some preconceptions derived from my background and from my reading of the literature, yet sought to remain open to new understandings that emerged whilst attempting to bracket out any preconceived notions I may have held of participants.

Finally, as the sole interpreter I recognise that data interpretation is by definition influenced by me (Joffe, 2012). Hence, following Vaismoradi et al. (2016), at certain points during the process of thematic analysis I chose to distance myself from the data for periods. Immersion in the data was obviously essential for coding and thematic development, but I recognised at times the value of taking a step back from it, revisiting my findings and interpretations with something of a fresh and self-critical eye.

1.4 Organisation of the thesis

I have structured this thesis across eight chapters.

In this first introductory chapter, I have given the background to and rationale for this study with reference to my own personal starting point as a business school academic, together with relevant aspects of the literature on critical thinking and assessment in higher education. I have further delineated the parameters of the study through research objectives, questions and the case setting. Finally, I have acknowledged the subjectivity of this research and set out my own reflexive position in response.

In Chapters 2 and 3, I present a two-part review of the literature base underpinning this study. In Chapter 2, I focus on how critical thinking in higher education may be defined, beginning with and expanding upon the consensus statement of the Delphi Project (see Facione, 1990), which sets out critical thinking in terms of skills and dispositions. In Chapter 3, I consider the difficulties of assessing for critical thinking in students' work, situated within a broader discussion of assessment practices in higher education, with attendant issues of assessment variation and ways in which institutions look to mitigate this.

In Chapter 4, I present my research philosophy, together with the methodology and methods I employed to develop answers to the research questions posed in Chapter 1. I provide brief descriptions of the study's participants, together with detailed accounts of data collection and analysis. I also consider the limitations of this research, attendant ethical considerations, and of how I have sought to maintain rigour with reference to suitable criteria.

In Chapters 5 and 6, I present the detailed findings of my analysis of the dataset, demonstrating the development of and support for the resultant themes.

In Chapter 7, I present a discussion of the findings from the preceding two chapters with reference to my research questions, RQ1-RQ3 inclusive, and the literature base In Chapters 2 and 3 together with any further reading channelled from the data analysis.

In Chapter 8, I present a brief overview of this thesis, offer a summation of the key findings in answer to my research questions, develop some theoretical and practical contributions, put forward suggestions for extension through further research, and conclude with some closing remarks.

2. Literature review (Part I) – critical thinking

2.1 Introduction

In the preceding introductory chapter, I questioned what we know of how academics conceptualise critical thinking. In context, critical thinking is a key goal for education according to the Davos World Economic Forum (2019). Higher education institutions place heavy emphasis on the development of this and other higher-order graduate skills, with critical thinking at the core of the intellectual mission of higher education globally (de la Fuente, 2009; Danvers, 2018). However, critical thinking is a contentious term that is subject to wide interpretation, is lacking in academic consensus, and which is then subject to multiple definitions (Moore, 2013; Wendt & Ase, 2015).

Evidently, the business of assessing students' work in universities is done year on year and for this to happen it stands to reason that if critical thinking forms part of the assessment then there must be a level of consensus of what this is. Internalised within academic discourses, academics absorb what critical thinking is into their practices so that they can recognise it when they see it despite the uncertainty and the apparent difficulties they have with defining and explaining what it is exactly (Atkinson, 1997; Fox, 1994; Gee, 2015). How business school academics in particular conceptualise critical thinking is an issue I investigate through this thesis with reference to RQ1. However, as a precursor, and for comparative purposes, I draw on relevant literature here in Chapter 2 to examine what critical thinking is with respect to higher education more broadly, developing a description of critical thinking as a practical concept for academics comprised primarily of skills and dispositions.

2.2 Defining critical thinking

The term ‘critical thinking’ is used pervasively in higher education. It features strongly in institutional rhetoric. Yet, as I have stated, there is no definitive definition, no precise specification, no clear conceptualisation, and no commonly held view of what critical thinking is, with many interpretations forcibly advocated. Unsurprisingly then, Moore (2013, p.519) called critical thinking “a contested notion”, Moon (2008, p.63) a “messy concept”, and Brookfield (1987, p.11) a concept that is “exhortatory, heady and conveniently vague”.

Critical thinking is a multi-dimensional concept (Liu, Mao, Frankel & Xu, 2016). At one level, critical thinking ‘as logicality’ encompasses the application of the rules of logic and reasoning in deconstructing arguments. Ennis (1993), for example, offers a detailed description that would link critical thinking with logic as a method for distinguishing correct from incorrect reasoning in arguments. At quite another level, critical thinking can embrace debates around critical pedagogy in education and critical citizenship in modern society; debates which I loosely categorise under ‘sociocultural views of critical thinking’. Barnett (1997), for example, argued that the purview of higher education should be the development of people who are critical beings, i.e. persons who, being self-aware and reflexive, can through critical reasoning make critical decisions and take critical actions that benefit society. What I have observed from looking across the literature are conceptual frameworks that present critical thinking in different ways by emphasizing certain aspects of critical thinking, for example philosophical ways of thinking including the application of logic and reasoning, the application of cognitive skills, attitudinal or dispositional requirements, and beyond to Barnett’s critical beings. I acknowledge the latter. However, whilst sociocultural views of critical thinking have value as part of a debated and comprehensive scholarly conceptualisation, they are of limited practical use as regards how academics actually operationalise critical thinking through assessment, with reference to the focus of this thesis. It follows that the literature considered here should centre on what Davies (2015, p.64) referred to as “the philosophical account of critical thinking”, comprised primarily of skills and dispositions qua argumentation and judgement.

2.2.1 The Delphi Project (1988-90)

Facione (1990) provides a key starting point and an authoritative base for any definitional discussion of critical thinking. Facione conducted a research project sponsored by the American Philosophical Association – the Delphi Project (1988-90) - that had as its remit the investigation and preparation of a framework for critical thinking. This aimed to compile a definition of critical thinking through cross-disciplinary discussions involving many international experts (Lincoln & Kearney, 2019). This resulted in a "consensus statement regarding critical thinking and the ideal critical thinker" (p.3), which reads as follows:

> "We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based. CT is essential as a tool of inquiry. As such, CT is a liberating force in education and a powerful resource in one's personal and civic life. While not synonymous with good thinking, CT is a pervasive and self-rectifying human phenomenon. The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. Thus, educating good critical thinkers means working towards this ideal. It combines developing CT skills with nurturing those dispositions which consistently yield useful insights and which are the basis of a rational and democratic society." (Facione, 1990, p.3)

This statement is rooted in the “critical thinking movement” of the 1970s and 1980s (Davies, 2015, p.42), which draws heavily from the disciplines of in particular philosophy and cognitive psychology. As a definition of critical thinking it is wordy and convoluted, which is symptomatic of the complexity of the subject, and Facione admitted that only a degree of consensus was reached with alternative lists proposed and at least one expert excluded by request from the consensus statement (Moore 2011a). Examining that consensus statement, I distinguish two core threads. Firstly, a cognitive skills outlook, i.e. the actions performed by critical thinkers pursuant to a task (qua argumentation and judgement formation), typically set out as cognitive skills, beginning with but not limited to the structured application of logic and reasoning (Rear, 2019; Swanwick et al., 2014). Secondly, a dispositions outlook, i.e. that critical thinkers possess the attitudes or dispositions necessary to motivate the exercise of said skills pursuant to critical thought (Davies, 2015; Halonen, 1995).

2.2.2 Cognitive skills outlook

2.2.2.1 Logic and reasoning

Critical thinking is often connected with the application of the formal structures of logic and reasoning vis-à-vis argumentation (Swanwick et al., 2014). Siegel (1985), for example, makes links between critical thinking and logic. Facione (2011, p.20) refers to a “zealous dedication to reason”. Fisher (2001, p.13) too connects critical thinking and reasoning, describing the former as “a kind of evaluative thinking” that is “particularly concerned with the quality of reasoning or argument…” Ennis (1962, p.86) originally defined critical thinking as “correct assessing of statements”, referring to the application of the rules of logic to arguments, although he later augmented this with reference to reflective judgement. Kurfiss (1988, p.iii) defined critical thinking as the ability “…to detect and avoid fallacious reasoning and to analyse deductive and inductive arguments”. Halpern & Sternberg (2020, p.3) also speak of “overcoming fallacies”, i.e. of identifying faulty reasoning in arguments. Mulnix (2012, pp.464-465) similarly defines critical thinking as “acquiring, developing, and exercising the skill of being able to grasp inferential connections holding between statements”, effectively reducing it to a process of logical reasoning.

For de Bono (1982), thinking more broadly and logic have long been associated because historically when students have been taught to think the focus has been on the formal rules of logic. However, whilst de Bono acknowledged logic as a useful tool for unpicking arguments, supported by Bernstein (1995, p.22) who deemed logic to be "a powerful guide for critiquing arguments", he also considered thinking to involve much more than the application of logic to arguments. Meyers (1986) would further claim that the ability to think critically is not necessarily improved by learning logic. Certainly, an argument can be entirely logical as constructed but this does not necessarily mean that it is sound (de Bono, 1983). Moreover, there is a sense of the critical thinker as autonomous, as someone who works through things to arrive at appropriately judged conclusions and justifiable decisions. This would seem incongruous with slavishly following the rules of logic. Nonetheless, several authors note that logic and reasoning are implicit to argumentation and hence, to critical thinking (see, for example, Bailin & Siegel, 2002; Erikson & Erikson, 2018; Moon, 2008; Pithers & Soden, 2000; Swanwick et al., 2014).

2.2.2.2 Skills of argumentation

More broadly than logic and reasoning, critical thinking is invariably characterised in terms of cognitive and interpretive skills (see, for example, Halpern, 2014), or 'mental processes' (Bailin, Case, Coombs & Daniels, 1999a), connected with analysing the arguments of others. Butler & Halpern (2020, p.152) describe critical thinking as "good thinking that is well reasoned and well supported with evidence". This would encompass *inter alia* "evaluating evidence, analysing arguments, inductive and deductive reasoning, identifying assumptions and hypotheses, drawing conclusions, extrapolating inferences and understanding implications" (Liu et al., 2016, p.678).

Critical thinking as argumentation involves being able to determine the validity of an argument as structured, assessing the extent to which its premises may be relied upon, whether its reasoning is fallacious or valid, whether its conclusions follow logically, etc. (Davies, 2015). This understanding of critical thinking as argumentation is ubiquitous for many, particularly so in higher education where argumentation is regarded as a fundamental skill (see, for example, Andrews, 2015; Davies, 2013, 2015; Mitchell, 1994; Mitchell & Riddle, 2000; Scott, 2000; Wingate 2012). This view of critical thinking demands that higher education students are able to understand, reconstruct, assess and evaluate arguments (Carrington et al., 2011; Harrell, 2011). These are the cognitive

tasks inherent to ‘argument analysis’ (Ennis, 1985a; Fisher & Scriven, 1997; Kuhn, 1991). In more detail, critical thinkers can recognise relationships, make suitable inferences and evaluate evidence; can track inconsistencies in reasoning; can pick apart arguments, root out central issues, question assumptions and assertions, synthesise information and interpret the validity of conclusions (Espey, 2018; Pascarella & Terenzini, 1991; Tsui, 2002). Critical thinking should also be evident in students constructing their own arguments (Moon, 2008). However, the lack of argument in students’ work, manifest in largely descriptive uncritical text, is a common complaint of academics (Du Boulay, 1999; Fox, 1994; Shafer, 2013).

Mayer and Goodchild (1995) narrowly construed critical thinking in terms of the ability to understand and evaluate arguments. This is arguably prevalent given, for example, the emphasis on the skills of argumentation as critical thinking in the study skills approach adopted by student self-help guides (see, for example, Cottrell, 2017). Skills are likely to be the easiest facets of critical thinking for students to grasp because they are familiar, residing within their existing cognitive frames of reference (Moon, 2008). Consequently, these self-help guides focus on deconstructing arguments as a developmental skill. Students should be able to make inferences from analysis and synthesis of relevant sources because the ability to analyse a range of material and formulate balanced and supported arguments is a staple of high grades, particularly in the social sciences and the humanities (Tynjala, 2008; Wendt & Ase, 2015; Guo & Shi, 2016). Students who are thinking critically are able to develop their own viewpoints, informed by their interrogation, analysis and evaluation of the ideas and arguments of others (Ramage et al., 2009; Jonassen & Kim, 2010; Luque, 2011; Hammer, 2017). However, the focus on skills development in self-help guides is largely with respect to studying the arguments of others, sequenced as a series of analytical steps – identify, analyse and evaluate the lines of reasoning, propositions, conclusions, evidence, etc. This provides students with a generic blueprint to follow, the result of which should offer evidence of critical thinking for academics. Agreeing with Moon (2008), some emphasis, although arguably not enough, is also placed on students constructing their own arguments. Nevertheless, the cognitive models of critical thinking proffered by such guides are popular and in all likelihood underpin much of how we engage with critical thinking in higher education because of its distillation into a series of analytical skills and steps. This is easier to teach, easier to learn and easier to evidence than engaging with critical thinking as a way of thinking or being. This approach may have its limits, simplifying from the multi-

dimensional concept that critical thinking is taking into consideration the literature as a whole. Nonetheless, it has been hugely influential – the multiple editions and volume of sales of such critical thinking guides would point to this – and is likely therefore to underpin students' behaviour and academics' expectations.

From the literature base the many skills of the critical thinker may be summarised as describing, understanding, questioning, interpreting, analysing, synthesising, evaluating, reasoning, inference, problem solving, self-regulation and judgement (see, for example, Bissell & lemons, 2006; Davies, 2015; Facione, 1990; Halonen, 1995; Halpern, 2014; Ikuenobe, 2001; Kek & Huijser, 2011; Nicholas & Raider-Roth, 2011; Paul & Elder, 2006; Rodriguez, 2009; Tapper, 2004).

2.2.2.3 Reflection and self-correction

Ennis emphasised the rational and reflective basis of the concept of critical thinking (Moore, 2013). Ennis's (1987, p.50) description of "reasonable, reflective thinking focused on deciding what to believe or do" is widely cited as articulating the nature of critical thinking (Lai, 2012). However, this is subject to the criticism that in its apparent simplicity the definition fails to capture the complexity of interpreting what is 'reasonable' and what is 'reflective' in application to 'thinking' (Leong, 2013). McPeck (1981) refers also to 'reflective scepticism', reflective being of relevance here, scepticism I will return to (see s.2.2.3). Connectedly I would add "Skilful, responsible thinking that facilitates good judgement" (Lipman, 1988, p.39); "thinking aimed at forming a judgement" (Bailin et al., 1999b, p.287); and "discerning or discriminating thought characterised by careful analysis and judgement" (Cooper & Patton, 2009, p.2). I note that these descriptions would seem to focus less on the mechanics and skills of argumentation and more on the reflective basis for forming judgements and making decisions. Such references to 'reflection' arguably serve to espouse a metacognitive approach to critical thinking.

Metacognition may be defined generally as thinking about thinking, or knowing what we know and what we do not know (Livingston, 1997; Mahdavi, 2014). Paul (1990, p.32) does not explicitly reference metacognition, although the inclusion of "the art of thinking about your thinking" within his list of facets of critical thinking could conceivably come within this. Elsewhere, it is suggested that to think critically we have to be self-regulatory, i.e. not just aware of our own thinking but actively monitoring and controlling our thinking (Facione, 1990; Griffith & Ruan, 2005; Hennessey, 1999; Martinez, 2006). Self-correction – being "brave enough to risk being wrong, and wise enough to realize that much can be learned from errors and failed solutions" (Nelson, 2005, p. xiv) – is, per Lipman (2003), essentially metacognitive and is said to be necessary for thinking to qualify as critical. We can learn and improve by being alive to potential deficiencies in our thinking through constantly testing our rationality against differing perspectives (Brookfield, 1987).

Some would argue that metacognition stands apart from critical thinking, whereas others, like Lipman (2003), see metacognition as important to developing an adequate understanding of critical thinking (see Davies, 2015; Halonen, 1995; Halpern, 1998; Jones, 2020; Kuhn, 1999; Kuhn & Dean, 2004; van Gelder, 2005; Willingham, 2008). Davies (2015), for example, argues that metacognition is integral to critical thinking as a necessary cognitive skill. Not a skill of argumentation *per se*, but one that is necessary for it. This, however, draws on only a limited body of research into cognitive development that specifically links metacognition or meta-knowing and critical thinking (see, for example, Kuhn, 1999; Olson & Astington, 1993). Moreover, Davies (2015) and Halonen (1995), who both advocate for metacognition as a cognitive skill for critical thinking, sought to build holistic conceptual models of critical thinking in higher education without having specific regard to its assessment. Metacognition is about self-assessment, not academics' assessment of students' work, as is the focus in this thesis. This fundamental difference obviates further consideration of metacognition as a claimed cognitive skill for critical thinking, at least in this thesis.

2.2.3 Dispositions outlook

Added to the cognitive skills outlook are dispositions. Davies (2015, p.55) refers to these as elements of 'propensity', meaning "an inclination or tendency to behave in a certain way." A person needs not only the skills to think critically but also the drive or propensity to use them (Ennis, 1987, 1996). Studies from education have shown that skills and dispositions relating to critical thinking are correlated (see, for example, Heijltjes, Van Gog, Leppink & Paas, 2014). Hence, students who think critically need to want to engage in critical thought (Bensley et al., 2016; Verburgh 2019; Wilson & Howitt, 2018). One can possess all the necessary skills for critical thinking but not apply them if disinclined to exercise those skills (Calma & Davies, 2020). The requisite dispositions or 'habits of mind' thus form precursors to the exercise of the cognitive skills pursuant to critical thought (Brown & Rutter, 2006; Ennis, 1985b; Facione, 1990). Here now the critical thinker is intrinsically motivated, possessing the right "attitudes and a sense of psychological readiness of the human being to be critical…defined [I suggest with reference to Facione, Sanchez, Facione & Gainen, 1995] as a constellation of attitudes, intellectual values and habits of mind" (Davies, 2015, p.55). Passmore (1967) and Facione (1992) both capture this under the term "critical spirit", manifest for Facione in "probing inquisitiveness" and "keenness of mind" (Facione, 1992, p.10). The literature offers several such dispositions inherent to critical thinking. These include open-mindedness, fair-mindedness, creativity, flexibility, curiosity, inquisitiveness, persistence, objectivity, a desire to be well-informed, prudence in decision-making and a readiness to self-correct where needed (see, for example, Bailin et al, 1999b; Calma & Davies, 2020; Ennis, 1985b; Facione, 1990; Facione et al., 1995; Halpern, 1998; Halpern & Sternberg, 2020; Paul, 1992).

Critical thinking in higher education is fundamentally about making reliable and persuasive judgements substantiated through sound assessment of evidence from various sources (Moon, 2008). However, critical thinking is evidently more complex and involved than the skills outlook would suggest, characterised not only by analysis and argumentative reasoning but also by authority, a willingness to doubt and hence a degree of independent thinking (Fisher, 2001; Harrell, 2011; Wendt & Ase, 2015). This is underpinned by a propensity for scepticism towards authorities (Moore, 2013; Rear, 2019). Students are expected to have the habit of mind to question and challenge, to examine alternatives, to strive for meanings and justifications (Moon, 2008). This

scepticism is controlled, putting it in a positive light in that 'critical' students are not wedded to their prior beliefs, nor do they accept information unquestioningly but exercise caution in absorbing knowledge and the ideas of others after carefully examining their reasoning, conclusions and evidence (Garside, 1996; Stanovich & West, 1997; Browne & Freeman, 2000). Linked with this, Paul (1981, p.2) differentiates "strong sense" from "weak sense" critical thinking. The latter refers to the process of applying the aforementioned cognitive skills to discover flawed reasoning, whereas the former acknowledges the complexity of situations, which invariably require choosing between different perspectives, and the importance of forming a judgement that is contextually situated. It is not just about being analytic therefore, although this is an important aspect of critical thinking, but also about possessing a "concern for truth" and being "open to new ideas" (Halpern and Sternberg, 2020, p.1). The critical thinker then is predisposed to "open-mindedness, fair-mindedness, independent-mindedness, an inquiring attitude…" (Lai, 2012, p.936). Being open to different interpretations, different perspectives, different approaches, enables the critical thinker to have greater appreciation of context and a deeper understanding of the position ultimately adopted (Browne and Freeman, 2000).

Similarly, for Brookfield (1987) critical thinking involves much more than the exercise of cognitive reasoning, and goes beyond dissecting the logic of arguments and weighing the opinions of others. It requires a recognition of our own beliefs, our own behaviours and the assumptions which underlie these which lead to our making seemingly rationale judgements. It is then an emotional as well as a cognitive concept. Emotions, both positive and negative, can motivate critical thinking (Brookfield, 1987; Halonen, 1995). As critical thinkers, we can experience "joy, release, relief, and exhilaration as we break through to new ways of looking at our…world" (Brookfield, 1987, p.7). If well managed, our emotions can help us to respond critically and contribute to critical thinking becoming our habitual response (Halonen, 1995). Emotion is conceivably inherent to several of the aforementioned dispositions connected with critical thinking by the Delphi Project consensus statement and the respective models of critical thinking in higher education of Davies (2015) and Halonen (1995). Emotion then is potentially as important to critical thinking as cognitive and affective elements (Davies, 2015; Halonen, 1995; Halonen, Smith & Dunn, 2008). However, this is not necessarily reflected in the literature, for which the key foci are cognitive skills and dispositions, with emotion given only very limited coverage.

Returning now to McPeck's (1981) 'reflective scepticism', I take this to mean firstly, being sceptical of conventional wisdom and dominant views (ala Mingers, 2000), and secondly, being open to alternative interpretations and perspectives (ala Paul, 1981), for the critical thinker tolerates ambiguity, appreciates individual differences and respects different viewpoints (see Bailin et al., 1999b; Facione, 1990). Invariably, thinking that is 'critical' is applied with deep engagement to complex matters that are likely subject to different viewpoints (Moon, 2008). Critical thinkers take nothing for granted, analysing assumptions and considering alternative perspectives, questioning their own assumptions and those of others and of received wisdom (Barnet & Bedau, 2010; Carrithers & Bean, 2008; Perkins, Jay & Tishman, 1994; Jones, 2009; Walker & Finney, 1999). For Brown and Rutter (2008), critical thinkers are not defensive and do not adopt a single-minded attitude but are instead open to possibilities. They have the courage to challenge orthodoxy but are not impulsive or precipitate in doing so, taking a prudent and thoroughly disciplined approach to thinking it all through and arriving at reasoned determinations (James et al., 2010: Jones, 2009; Mingers, 2000).

Finally, creativity also has a dispositional role in critical thinking (Bailin et al., 1996b; Moon, 2008). Thinking critically is said to involve thinking creatively (Halpern, 2014). Being creative in the higher education context involves "producing something new" (Dumitru, 2019, p.870). It can involve exercising ingenuity and innovation (Anastasiadou & Dimitriadou, 2011) in doing something that is "novel and adaptive" (Brodin, 2016, p.972, drawing on Amabile, 1996 and Sternberg & Lubart, 1999), but which is also "relevant (valuable)" (Bennich-Bjorkmann, 1997, p.25), denoting for Anastasiadou and Dimitriadou (2011) the importance of creating new knowledge. Novelty can come from the context, participants and subject of research (Dumitru, 2019). This is likely to involve some imagination as well as taking on some risk (Sternberg & Kaufman, 2010). Creativity is thus said to embody originality (Brodin, 2016; Dumitru, 2019; Jackson & Shaw, 2006; Lovitts, 2008; Winter, Griffiths & Green, 2000). Originality is in turn closely associated with critical thinking (Simpkins, 1987; Holbrook et al., 2007). I have said that students in higher education are expected to move beyond passive reception of that which is given as known through to active engagement in critical thinking. This extends even to creation of new knowledge (Bok, 2006). Moore (2013) too highlighted the importance, as far as participating academics in his study were concerned, of students actually going beyond challenging the ideas of others to demonstrate some originality of

thought in producing their own ideas, of contributing to knowledge in some way, howsoever modest.

2.2.4 The generalist versus specifist debate

What critical thinking is, or how critical thinking is defined, is one substantial debate in the literature that I have sought to examine in this chapter. Another substantial debate, in the context of how critical thinking might best be taught and developed in students, is the debate highlighted by Davies (2013) between what he refers to as the 'generalists' and the 'specifists', i.e. whether critical thinking is a generic skill or a disciplinary specific one. On the one hand, the generalists subscribe to the view that critical thinking is largely if not wholly a generic ability which is transferable across disciplines and contexts (Ennis, 1962), and which can thus be taught and learned independently of disciplines. Whereas the specifists hold that critical thinking is a discipline-specific ability which is not so transferable (McPeck, 1981) and which cannot be taught and learned absent the relevant discipline. A "long-standing controversy", this debate has continued more recently between Davies (2006, 2011, 2013), who advocates the generalist position, and Moore (2004, 2011b), who likewise supports the specifist position (Golding, 2011, p.360). I acknowledge this debate but need not enter into it here because as I indicated the discourse underpins a broader pedagogic debate about how critical thinking may be taught, rather than how critical thinking may be assessed.

Nonetheless I was conscious of this debate when forming this research and in structuring my participant sample (refer to s.4.4.3 Participant sampling) with a view to potentially exploring this issue of critical thinking as generalist or specifist across the different subject disciplines of a business school. Ultimately, I did not place this in focus through the study's research questions, concentrating therein on conceptualising critical thinking and examining how it is operationalised through assessment. Hence, I did not take this issue forward in this thesis. However, this is something I would hope to revisit post thesis (refer to s.8.5 Further research for more information).

2.3 Conclusion

Critical thinking is defined in the literature, albeit in multiple ways, but many of which converge on the same basic awareness of critical thinking as an amalgamation of skills and dispositions (Halpern & Sternberg, 2020). The Delphi Project, for example, did just that and the project's consensus statement serves as an authoritative starting point for a definitional journey into critical thinking in higher education.

Subsuming logic and reasoning, the skills outlook characterises critical thinking in terms of cognitive and interpretive skills connected with argumentation (Moon, 2008), i.e. skills of interpretation, analysis, evaluation, inference and metacognition, pursuant to judgement formation and decision-making. This corresponds to the first part of the Delphi Project consensus statement. In higher education, this is about students developing and applying their cognitive skills to deconstruct the arguments of others and to make reasoned arguments of their own, forming substantiated judgements or decisions. This skillset is of course of value across disciplines, particularly so in business contexts in which rapid complex decision making is often a necessity of working life.

Critical thinkers are not solely reliant upon cognitive skills pursuant to critical thinking. The literature offers several relevant dispositions, e.g. open-mindedness, fair-mindedness, creativity, curiosity, flexibility, self-correction, etc., noting that skills and dispositions coordinate for critical thinkers to achieve purposeful reflection in reaching judgements and making decisions (Payan-Carreira, Cruz, Papathanasiou, Fradelos & Jiang, 2019). One needs the skills but also the propensity to use them to engage in critical thought.

In this chapter, the first of a two-part literature review, I have focussed on what critical thinking is as a practical concept with respect to higher education, linking with RQ1. In the following chapter, Chapter 3, I focus on how assessment practices accommodate critical thinking, why variation occurs by and across assessors and what institutional mechanisms exist to mitigate this, linking with RQ2 and RQ3.

3. Literature review (Part II) – critical thinking and assessment practices

3.1 Introduction

In the preceding chapter, I sought to define what critical thinking is with regard to higher education, primarily as an amalgamation of skills and dispositions. This relates to RQ1: How do business school academics conceptualise critical thinking?

In this chapter, the second of a two-part literature review, I extend my focus to how critical thinking may be assessed in higher education, situated within a broader discussion of higher education assessment practices. This relates to RQ2: How are business school academics' conceptualisations of critical thinking operationalised through assessment practices?

Assessing a student's piece of work means attaching to it a grade, i.e. a level of achievement in a currency that is commonly understood as evidencing and communicating the academic capabilities of students, such as a percentage mark or other numerical or qualitative classifier (Sadler, 2009). In higher education, critical thinking invariably forms part of this assessment but as to how exactly and to what extent are unclear. I consider these points in relation to higher education assessment practices. I begin by problematizing critical thinking and assessment in higher education, distinguishing standardised tests of critical thinking from assessing for critical thinking in essays and dissertations, and highlighting the problems of and reasons for grading variation. Variation notwithstanding, it seems that ultimately in reaching examination boards there is apparent agreement between assessors. I put forward two explanations for this. Firstly, I make the case for models of epistemological development, examining in particular those of Perry (1970) and Baxter Magolda (1992), as potentially reflective of academics' expectations of students with respect to critical thinking as assessed. Secondly, I consider the part played by key institutional mechanisms put in place to achieve consistency and hence mitigate variation, namely criterion-referenced assessment, internal moderation and external examination. This all links with RQ3: What are the challenges to achieving consistency when assessing for critical thinking in undergraduate students' dissertations?

3.2 Assessment of critical thinking problematized

If higher education institutions are nurturing critical thinkers then this should be apparent from and demonstrated through assessment of learning outcomes (Arum & Roksa, 2008; Klein et al., 2007; Lee at al., 2015). Assessment can take many forms but higher education assessment practices are underpinned by a culture of testing (Medland, 2016). One way of evaluating critical thinking is through standardised tests. There is now an industry to this with several tests devised for this purpose. Evidently there is some value in these tests which, built on the critical thinking taxonomies of Ennis, Facione, etc., seek to evaluate skills of analysis, deduction, inference, evaluation, reasoning and so on (Rear, 2019). Their proliferation would suggest that assessment of critical thinking whilst difficult is nonetheless feasible (Ennis, 1993). However, this undoubtedly remains problematic (Black, 2012; Tiruneh, De Cock, Weldeslassie, Elen & Janssen, 2017). Ennis (1993), for example, highlighted several 'traps' for the unwary tester of critical thinking that serve to interfere with validity. Rear (2019, p.668) describes such tests as "limited in their extent", based on restricted conceptions of critical thinking, assuming only correct/incorrect answers, which in itself appears uncritical, and "inauthentic" in being divorced from "real-world examples". Of particular concern is the emphasis given to standardised tests as quantitative assessment instruments, with greater focus on qualitative aspects called for when assessing critical thinking (Puig, Blanco-Anaya, Bargiela & Crujeiras-Pérez, 2019). Measurement of learning forms the dominant discourse of higher education assessment culture - we measure outcomes and we certificate learning (Boud, 2007). However, Yorke (2011) argues that putting the emphasis on 'measurement', whether implicit or explicit, is inappropriate because grades are not true measures in the scientific sense. In any event, the assessment of critical thinking in higher education does not typically take the form of standardised testing given that there is often no correct answer or single way of demonstrating the achievement of complex learning outcomes. Assessment of higher-order tasks such as critical thinking is an exercise in quasi-measurement involving complex interpretive judgement over scientific measurement, which makes it challenging.

Students who achieve higher grades necessarily demonstrate greater critical thinking and achieve better grades than lower academic achievers (Dwyer et al., 2014; James et al., 2002; Stassen et al., 2011). However, whilst assessment is one of the most common duties performed by academics, it is also one of the most difficult duties to perform well (Whalley, 2016). Traditional signature assessment tasks leading to single rather than multiple outcome measures are insufficient to assess the complexity of critical thinking (Spicer & Hanks, 1995). Hence, Academics possess significant autonomy concerning assessment, and where complex work is involved student responses are generally open, diverse and given considerable latitude (Bloxham, 2009; Bloxham et al., 2011).

Essays are set as a means of fostering analytical thinking and independence in students and are commonly employed in assessing undergraduate students for complex learning aspects such as critical thinking (Wendt & Ase, 2015). Argumentative essays, a genre associated with higher order thinking, are particularly valued as mechanisms by which critical thinking can be assessed (see, for example, Andrews, 2003; Hammer, 2017; MacLellan, 2004; Prosser & Webb 1994; Scott, 2000; Smith, Campbell & Brooker, 1999). This is especially observable in so-called 'soft disciplines', i.e. non-science subjects, to include business-related subjects (Ylonen, Gillespie & Green, 2018). Essays are thought able to offer insights into students' critical thinking because they demand careful analysis and critique of the views of others, evaluating their arguments and providing supporting evidence (Leong, 2013; Schendel & Tolmie, 2017; Tapper, 2004). However, students may tackle an essay question in different but equally effective ways. This necessarily lends to subjectivity, causing problems for reliability between assessors (Baume, Yorke & Coffey, 2004; Bloxham 2009; Yorke, 2011). Students also find it difficult to produce essays that show suitable levels of critical thinking (Hammer & Griffiths, 2015). Assessing this consistently is thus no easy task. Of course any claims to comparability and consistency make assumptions that academics hold shared conceptions of academic standards as applied to students' work, assumptions which research has questioned severely (Bloxham, 2009; Bloxham et al., 2016a).

That higher education assessment is inconsistent and unreliable is well observed, researched and reported (see, for example, Bloxham et al., 2016a; Ecclestone, 2001; O'Hagan & Wigglesworth, 2015; Yorke, 2008, 2011). With undergraduate numbers rising leading to larger assessment teams inconsistencies have become more apparent (Mulryan-Kyne, 2010; O'Hagan & Wigglesworth, 2015). Even in the context of large-scale standardised tests, studies have shown that assessor variability is a concern (see, for example, Barkaoui, 2010; Lumley, 2005; Weigle, 2002; Wolfe, Kao & Ranney, 1998). Considerably more variation is observed when it comes to assessing complex higher-order thinking through essays, for example, because unlike standardised tests these are open to individual interpretation and resist direct measurement (Bloxham et al., 2016a; O'Connell et al., 2016). Variation is similarly observed, with particular reference to this thesis, with respect to the assessment of dissertations (see, for example, Bettany-Saltikov et al., 2009; Pathirage, Haigh, Amaratunga & Baldry, 2007).

3.2.1 Grading variation

Assessor reliability is observably low for essays in cases outwith pure knowledge recall (Elton & Johnson, 2002). O'Hagan and Wigglesworth (2015) found distinct variability in grades awarded to the same essay by different assessors, in line with previous studies of a similar nature (see, for example, Laming, 1990; Read, Francis & Robson, 2005). It seems that complex qualitative performances simply cannot be reduced to that which can be measured with certainty or reliability (Bloxham, 2009; Knight & Yorke, 2003; Yorke 2008). The open nature of an essay or a dissertation renders it more difficult to achieve reliability compared with more clearly defined closed assessment tasks such as multiple-choice tests (Huot, 1990; O'Hagan & Wigglesworth, 2015). Essays that involve the human world, as opposed to the world of the natural sciences, necessarily rely more heavily on assessors' subjective judgments (Knight, 2006). However, studies in the field of cognition have shown that characteristics inherent to complex decision-making cause inconsistencies in academic judgement (Brooks, 2012). Healey (2013), referencing Webster et al. (2001), suggests that variation is often higher for dissertations than seen elsewhere, pointing to the worth of double marking processes as a check on this (see s.3.3.2).

Why variation occurs is undoubtedly a complex matter (George-Williams, Carroll, Ziebell, Thompson, & Overton, 2019). To begin with, mistakes or ineptitude cannot be ruled out (Dalziel, 1998; Heywood, 2000). Research has also shown that variation can be ascribed to differences in implicit beliefs, i.e. what each academic individually considers of importance or worthy of particular attention when judging the academic worth of a piece of work (Hunter & Docherty, 2011). Academics have different backgrounds, knowledge, values, expertise, experience, etc. and may as a result differ in the importance each attaches subjectively to particular aspects or qualities of assessed work (Bloxham, 2009; Bloxham, Hudson, den Outer & Price, 2015; Bloxham et al., 2016a; O'Hagan & Wigglesworth, 2015; Read et al., 2005; Smith & Coombe, 2006). That assessors operate from such different reference points is perhaps also reflective of different levels of socialisation into disciplinary and/or local assessment communities (Johnston, 2004; Yorke, 2011; Yorke, Bridges & Woolf, 2000). Assessors may, for example, not sufficiently understand the published assessment outcomes; or have differing conceptions of what assessment actually involves; or differing expectations of what constitutes achievement at various levels (Baume et al., 2004; Ekstrom & Villegas, 1994; Grainger, Purnell & Zipf, 2008; Hand & Clewes, 2000; Yorke, 2011; Yorke et al., 2000). Any of these factors may lead to variation.

Despite these observed issues, academics are considered expert assessors (Bloxham, 2009). Due to their knowledge, formed from their background, education, expertise, experience and understanding of the standards of their discipline borne of disciplinary socialisation, academics are appropriately positioned to make sound and reliable expert assessment judgements (Bloxham et al., 2011; Ecclestone, 2001; Sadler, 2013). This is implicit to non-standardised elements of higher education assessment. Expert assessors reach a point through experience where they acquire “practical mastery” of assessment systems through which they are able to judge students’ work, albeit often struggling to articulate the basis of that judgement (Bloxham, et al., 2011, p.658). The more proficient expert assessor is able to intuitively judge and explain the derivation of that judgement with reference to assessment criteria (Dreyfus & Dreyfus, 2005). In the main, people become better at something with practice and experience. However, experience of assessment is no guarantee of consistency. With respect to the consistent application of assessment criteria, for example, studies have found no apparent difference between novice and experienced assessors, notwithstanding the apparent greater propensity of

novice assessors to be more rule governed (Bloxham et al., 2016a; Ecclestone, 2001; Price, 2005).

As learning increases in complexity, and with it chosen modes of assessment, the more assessors rely on what Eisner (1985) referred to as ‘connoisseurship’ over measurement in judging students’ work (Bloxham, 2009). Academics rely less and less with experience on the application of artefacts such as predefined assessment criteria (Jansson, Carle, Gunnarsson, & Ekbrand, 2019; Sadler, 2005; Yorke, 2011). Academics as connoisseurs rely on their innate judgement having cultivated mental assessment models that they apply irrespective of assessment criteria (Ecclestone, 2001). This, however, tends to be unconscious, intuitive and tacit, making it elusive and its articulation and discernment troublesome (Golding, 2019; O’Donovan, Price & Rust, 2004). Moreover, expertise is no guarantee of quality. Expert subjective judgements can still be “poorly based, erratic and unreliable” (Sadler, 2013, p.14). Given these factors, it seems unlikely that assessors relying upon individual connoisseurship will give similar judgements, particularly of learning outcomes such as critical thinking that resist quantification. However, research has shown that development of shared understandings can work over time to calibrate judgements and reduce variability within teams (Adie, Lloyd & Beutel, 2013; Beutel et al., 2017; Elwood & Klenwoski, 2002; Sadler, 2011, 2013). Hence, less variability would be anticipated amongst assessors who frequently assess together. If connoisseurship is located within a local assessment community then the more experience an assessor has within that community the more likely it is that assessors’ judgements will align more closely.

Notwithstanding the issues highlighted here, and the evidence pointing to variation, the reality is that everything settles down in time for grades to be processed through examination boards. There is a tendency towards agreement with the grades submitted to examination boards outwardly agreed between assessors. In practice, this means that a second assessor has reviewed a sample of the grades awarded by the first assessor and professed agreement with these. Alternatively, that the second assessor has re-assessed the sample, the respective grades of first and second assessors are compared and final grades agreed. Local consensus is generally reached therefore, premised on academics having similar views of what is required of students’ work for assessment purposes, including then with respect to critical thinking.

I put forward two possible explanations for this. The first revolves around models of epistemological development in higher education. Knowledge is an important aspect of critical thinking in higher education as assessed. Subject knowledge, of course, because critical thinking requires a deep knowledge of the subject (Halpern & Sternberg, 2020). Also epistemological development, which I focus on in more detail in s.3.3, because how students perceive knowledge is important to their critical thinking abilities (Moon, 2008), and I suggest to the demonstration of such abilities through assessed work and academics' expectations of this. However, I am conscious that whilst academics may claim to agree grades in advance of examination boards research has questioned whether they do so for the same reasons, and there are apparently issues of shared interpretation of standards that work against consistency (Bloxham et al., 2015; Bloxham et al., 2016a; Pathirage et al., 2007). Appreciably, if the thinking that underpins assessment is individually tacit, unarticulated and unshared, then assessors will not develop the necessary shared understandings (Golding, 2019). This is where institutional mitigating mechanisms come into play. I offer these as a second explanation because they purportedly help to develop understandings and facilitate agreement between assessors. I cover this in more detail in s.3.4.

3.3 Academic expectations – the case for epistemological development

When seeking to explain apparent agreement between assessors I extended my literature search along various lines. One of these brought me to the concept of epistemological development. Empirical research has long linked epistemological development and critical thinking (see, for example, Brabeck, 1983; King & Kitchener, 1994; Kitchener & King, 1981; Kuhn, 1991, 1992). Kuhn argues that epistemological knowing is crucial to critical thinking. This rests on "an individual's broader understanding of knowledge and knowing" (Kuhn, 1999, p.18). Similarly, for Battersby (1989), also King and Kitchener (2004), both referencing McPeck (1981) and Siegel (1985), critical thinking has epistemic cognition at its theoretical foundation. Models of epistemological development in higher education – focussing in particular on Perry (1970) and Baxter Magolda (1992) – propose that students progress through several development stages in how they perceive knowledge. This is relevant because a person's epistemological development and their capacity for critical thinking are thought to be related (Kuhn, 1999; Moon, 2008). More recently, some education based studies of critical thinking – see, for example, Mingers (2000), Phillips and Bond (2004), Chen and Rattray (2017) – have findings that would seem to draw on epistemological

development, and in particular Baxter Magolda's (1992) model. Chen and Rattray (2017) actually included epistemological development, with reference to Baxter Magolda's (1992) model, in their working definition of critical thinking in higher education. I regard these papers as tangential to this thesis because their focus is invariably pedagogical, i.e. on the development of critical thinking in students rather than assessing for it. However, I reason that these models of epistemological development, built with reference to higher education students, could resonate for assessment in higher education because their descriptions of knowledge and knowing at progressive levels are potentially reflective of academics' expectations of what is and is not critical re students' thinking as demonstrated through their assessed work.

Baxter Magolda's (1992) Epistemological Reflection Model builds on the prior work and modelling of Perry (1970). The two provide comparable portrayals and common paths of progression in ways of knowing (West, 2004). Perry (1970) postulated that students move through four broad epistemological stages – dualism, multiplism, relativism and commitment to relativism – segregated further into nine progressively more sophisticated positions, effectively moving from absolute to interpretive knowledge stances (Brownlee, 2004). Baxter Magolda's (1992) Epistemological Reflection Model similarly has four domains of awareness, described as "ways of knowing" (p.29). The four domains – 'absolute knowing', 'transitional knowing', 'independent knowing' and 'contextual knowing' – are sequenced as a continuum representing growth in epistemic positions (Ostorga, 2006).

At the lower end of Perry's scale (dualist) are students who hold absolute interpretations of the world. Things are black or white, right or wrong, good or bad (Carroll, 2007). Truth can be known, facts are taken as read, and learners are passive seeking certainty and absorbing knowledge from authority figures or experts such as teachers whose role is to communicate the truth (Brownlee, 2004; Carroll, 2007; Hofer, 2001). Similarly, in Baxter Magolda's first domain (absolute knowing) knowledge is viewed as certain. Learning is gaining knowledge from experts who "have all the answers" (Hofer & Pintrich, 1997, p.98). Experts know what is true and offer certainty in what they say (Baxter Magolda, 2002). The student accepts and absorbs the knowledge provided in the certainty that it is expert knowledge (Moon, 2008). As I see it, this is not indicative of students being 'critical'.

In Baxter Magolda's second domain (transitional knowing) the learner's perception of knowledge as certain is weakened. Some knowledge remains certain and some now is uncertain (Baxter Magolda, 1992; Carroll, 2007). Faced with areas of uncertainty, transitional knowers work on their understanding of the knowledge rather than simply looking to acquire it (Baxter Magolda, 2002). Moving through Perry's positions, students become "more relativistic in their point of view" (Carroll, 2007, p.214). They gain awareness and recognition of multiplicity, acknowledging that experts can hold different views, that there is not necessarily a right or a wrong, that not everything can be known with certainty (Brownlee, 2004). There is evaluative recognition that multiple perspectives and conflicting viewpoints can exist and have validity, yet not all views are equally valid and some are better than others (Hofer, 2001). Students move towards forming and justifying their own opinions in the midst of uncertainty, opposing viewpoints and alternative frames of reference (Moon, 2008). Similarly, students in Baxter Magolda's third domain (independent knowing) recognise that knowledge is largely tentative and uncertain, that individuals carry their own beliefs and have the right to hold their own views, and that experts are not the only source of knowledge (Brownlee, 2004; Carroll, 2007; Hofer & Pintrich, 1997; Moon, 2008). Independent knowers are open to multiple perspectives (Baxter Magolda, 2002). They may not challenge them as such but will choose between them (Baxter Magolda, 1992). They think for themselves, now generating their own perspectives and holding these as equally valid (Baxter Magolda, 1992; Hofer & Pintrich, 1997).

Fuller epistemological development in Perry's model would categorise students in a relativist position. Here, students see knowledge as not absolute but actively constructed (Hofer & Pintrich, 1997; King & Kitchener, 1994). Truth is seen as relative to contextualised individual interpretations, and a point of view is developed consistent with the justification for it (Brownlee, 2004; Carroll, 2007). Similarly, in Baxter Magolda's fourth and most sophisticated domain (contextual knowing) knowledge is seen as constructed and understood in its context, legitimised and evidenced by that context (Baxter Magolda, 2002; Evans, Forney, Guido, Patton & Renn, 2010). Contextual knowers are highly analytical and highly critical persons who sift multiple perspectives in constructing their own perspectives judged in context and on the evidence (Baxter Magolda, 2002; Carroll, 2007; Evans et al., 2010; Hofer & Pintrich, 1997; Moon, 2008; Ostorga, 2006; West, 2004). They consider issues from all sides and with regard to all aspects and perspectives both expert (objective) and personal (relational) (Brownlee,

2004). If we think of knowledge as being absolute then it is unquestionable and accepted as just so and no critical thought can be applied to it (Kember, 2001; Kuhn, 1999). However, if we see knowledge as constructed, are alert to the complexity of that construction and open to questioning it, then critical thought can ensue. As I see it, this is far more indicative of students being 'critical'.

A key point to take from these models is that a student's capacity for critical thinking is likely to be refined as they progress through the stages of epistemological development. It is anticipated then that new undergraduates will reside at the absolutist end of the scale (Baxter Magolda, 2002). Their capacity to think critically is necessarily limited because their epistemological beliefs are relatively undeveloped at this stage. Also generally anticipated is that students will progress gradually away from an absolutist position, moving towards a position that recognises knowledge as constructed and relative. Their capacity for critical thinking grows as their epistemological beliefs develop (Moon, 2008). I see this as relevant to this thesis, with the emphasis on business school academics' views on assessing for critical thinking in final year undergraduate students' dissertations and their expectations of students at that level.

The models of Perry (1970) and Baxter Magolda (1992), together with related works (see Belenky, Clinchy, Goldberger & Tarule, 1986; King & Kitchener, 1994; and Kuhn, 1991) evidently view personal epistemology in terms of staged progression. Individuals who move through the stages develop in sequence seeing knowledge more and more as constructed and reflecting a growing ability to manage both objective and subjective parts (Hofer, 2004). In all these frameworks there is a trajectory envisaged, a progression from absolutism to relativism. I note too that they occupy a significant portion of the literature relating to personal epistemology (Hofer, 2001). However, at this point I should acknowledge that these models do not constitute a singular accepted view of epistemological development as concept and as pedagogic practice. There is evidently much debate around this with alternative views proffered and a significant challenge coming from the concept of 'epistemological beliefs' proposed by Schommer (O'Donovan, 2017). Schommer (1990), in contrast to staged development models, proposed that personal epistemology should be considered more a collection of beliefs about knowledge and learning that can operate together but which exist more or less independently. Hence, they do not exist in an integrated, sequential framework (Hofer, 2004; Hofer & Pintrich, 1997; O'Donovan, 2017).

Schommer, building in particular on the work of Perry, engaged in a series of linked studies (see Schommer 1990, 1993, 1994; Schommer & Walker, 1995). These resulted in five constituent epistemological beliefs, each cast as continua in themselves, and each relatively independent, challenging then the notion that personal epistemology develops in fixed, sequential stages (Hofer & Pintrich, 1997). These beliefs are (1) Source of knowledge ("From knowledge is handed down by omniscient authority to knowledge is reasoned out through objective and subjective means"; (2) Certainty of knowledge ("From knowledge is absolute to knowledge is constantly evolving"; (3) Organisation of knowledge ("From knowledge is compartmentalised to knowledge is highly integrated and interwoven"); (4) Control of learning ("From ability to learn is genetically predetermined to ability to learn is acquired through experience"; and (5) Speed of learning ("From learning is quick or not-at-all to learning is a gradual process" (Schommer, 1994, p.301). As for dimensions in the staged development models, each of Schommer's epistemological beliefs is relevant to critical thinking and its development (Moon, 2008). Given, however, the claimed independence of these beliefs, it seems that students as learners (and as critical thinkers) could be sophisticated in some beliefs but less so in others (Hofer & Pintrich, 1997). Evidently also for Schommer (1994, p.302), "epistemological beliefs do not necessarily develop in synchrony."

Regardless of the modelled approach, development (e.g. Perry, 1970 and Baxter Magolda, 1992) or beliefs (Schommer, 1994), personal epistemology is portrayed as consisting of specified dimensions that encompass individual ways of knowing. The key difference is whether these dimensions are seen to evolve in sequenced progressions, or independently and not 'in synchrony' as Schommer put it. I note that Hofer and Pintrich (1997) questioned the reasoning and methodology underpinning Schommer's work, indicating also that the weight of evidence in the literature is on the development side, with Schommer's counter view arguably in need of more evidence to establish its veracity. That said, they did acknowledge the usefulness of having such challenges to existing views of personal epistemology (Moon, 2008). I agree with this. However, whilst acknowledging Schommer, the sense I have from my reading of discussions with participants is of academics' expectations aligning with staged development in connection with critical thinking (I consider this further in s.7.2.4.1 Epistemological development).

A further key point then is that these models may speak to what is and is not 'critical' in students' work, reflecting academics' expectations. This is of particular relevance to assessment in higher education. Critical thinking in higher education is fundamentally about making reliable and persuasive judgements substantiated through sound assessment of evidence from various sources (Moon, 2008). Academics expect work from students which shows independent research, is analytical and which argues positions effectively, yet lament the standard of many submissions often labelled merely 'descriptive' or 'opinionated' (Hounsell, 2005 ; MacLellan, 2004; Olsen & Statham, 2005; Wendt & Ase, 2015). Students as critical thinkers should be able to make inferences from analysis and synthesis of relevant sources because the ability to analyse a range of material and formulate balanced and supported arguments is a staple of high grades, particularly in the social sciences and the humanities (Guo & Shi, 2016; Wendt & Ase, 2015). Students do not simply accumulate knowledge but must work to overcome resistance and uncover that which is hidden (James et al., 2010). Ultimately, they move beyond deconstructing the arguments of others to find patterns and make meaningful connections in developing their own arguments, reasoned appropriately to explain and persuade others (Buskist & Irons, 2008; Jones, 2004; Lundquist, 1999). They are able to develop their own evidence based viewpoints, informed by their interrogation, analysis and evaluation of the ideas and arguments of others (Hammer, 2017; Jonassen & Kim, 2010; Luque, 2011; Ramage, Callaway, Clary-Lemon, & Waggoner, 2009). This would place them in the more developed stages of the models of Perry (1970) and Baxter Magolda (1992), with greater capacity for critical thinking.

3.4 Academic expectations - institutional mitigating mechanisms

Criterion referenced assessment practices and systems of internal moderation and external examination are considered to have important roles in mitigating the issues raised in the preceding sections leading to assessment variation.

3.4.1 Criterion referenced assessment practices

Academic standards in UK higher education are outcome focussed, i.e. defined in terms of achievement levels that students must attain for an award (Bloxham et al., 2015). Achievement is demonstrated by performance enacted through assessment (Alderman, 2009). Outcomes-based approaches to assessment are invariably based on criterion

referencing. This developed from the work of Glaser and Klaus (1962) who first coined the term ‘criterion referenced measurement’, now firmly established in the assessment lexicon. Although Wiliam (1996) claimed this is an inadequate regime for achieving an authentic evaluation of performance, others are supportive. Gipps (1994), for example, argued that conscious inclusion of assessment criteria that cover higher-order cognitive skills serves to place the spotlight on this ensuring adequate attention is given. Criterion referenced based assessment is also said to be more objective and therefore a more ethical assessment practice (Sadler, 2009).

I stated in Chapter 1 that higher education assessment practices are located within an agenda of institutional accountability underpinned by quality assurance frameworks (Bloxham, 2009; Medland, 2019). This accountability paradigm assumes that student work is most appropriately assessed with reference to criteria (Bloxham et al., 2011). This is now common practice across the tertiary sector (Grainger et al., 2008; Lok, McNaught & Young, 2016). Publication and consistent application of assessment criteria are pivotal to the transparency required, outwardly serving to obviate academics’ prejudices and promulgate shared understandings between academics, and between academics and students (Andrade & Du, 2005; Bloxham et al., 2016a; Menendez-Varela & Gregori-Giralt, 2016; Venning & Buisman-Pijlman, 2013). Such efforts build confidence, confidence that is needed by academics and greatly appreciated by students (Jonsson, 2014). However, this confidence may be more assumed than real (Bloxham, 2009). Certainly, publication of assessment criteria is only part of the story. If criteria are not actually read and absorbed then their publication is seemingly wasted.

Each student’s work is judged individually against the assessment criteria, in isolation from his or her peers, to minimise the effects of personal judgement and increase assessment consistency and reliability (Grainger et al., 2008; Lok et al., 2016). However, views on whether this works in practice are mixed. Assessment reliability has been shown to improve from the use of criteria, if only to a small extent (see, for example, Jonsson & Svingby, 2007; Peeters, Schmude & Steinmiller, 2014). Criteria provide a supportive crutch to lean on, particularly for novice assessors, but are relied upon less and less with experience (Jansson et al., 2019; Crisp, 2008; Yorke, 2011). Other studies highlight much assessor discrepancy in the application of assessment criteria to determine grades (see, for example, Baume et al., 2004; Price, 2005; Yorke, 2008).

Outcomes with relatively low-level cognitive demand generally have clear definition and are simple to measure. These, however, are not generally the preserve of higher education, where outcomes such as critical thinking that are more cognitively demanding are not easily defined or measured (Lok et al., 2016). It is possible to construct criteria for this but very difficult to interpret them consistently, particularly absent from the relevant context (Neil, Wadley & Phinn, 1999; Woolf, 2004). Criteria are built from words, and words need a framework to imbue understanding, a framework which is necessarily contextual and situation-specific (Bloxham, 2009; Yorke, 2011). It is thus questionable whether complex learning can be adequately described in assessment criteria to avoid varying interpretations (Bloxham, 2009). In Grainger et al. (2008), for example, a group of assessors were shown to draw on the same overarching criteria but differences in interpretation led to variance in grades.

Words often used in assessment criteria which purport to denote levels of performance for assessment purposes such as ‘good’, ‘excellent’, ‘comprehensive’ or ‘adequate’ are relative terms that are invariably open to subjective interpretation (Grainger et al., 2008). Criteria that are loosely defined or too subjective are obviously difficult to interpret (O’Donovan et al., 2004). Constructing clear, concrete assessment criteria and scaling different standards of achievement or outcomes against these is evidently challenging (Lok et al., 2016). So criteria need to be clearly defined and contextualised for assessors to appropriately interpret them (Bird & Yucel, 2013; O’Donovan et al., 2004). Still, interpretations are rarely incontestable (Bloxham et al., 2016a).

Adding to the confusion here, in this context of criterion-referenced or ‘criteria-based’ (Sadler, 2005) assessment practices, it seems that ‘criteria’ and ‘standards’ are often confused and apparently used interchangeably (Bloxham, et al., 2011; Sadler, 1987, 2005). Sadler (2014, p.275), however, would refute this, viewing the two as distinct, related certainly but clearly distinguished, i.e. “A *criterion* is a property or quality used in appraising student responses to assessment tasks, whereas a *standard* is a minimum achievement level used as a reference point when judging the quality of a student’s work.” Criteria underpin standards (Sadler, 2014). Standards provide a “framework for making categorical attributions”, enabling statements about the quality of a student’s work or their level of achievement in isolation from the work of others (Sadler, 1987, pp.195-196).

For Bloxham et al (2011), Sadler's 'standards' are thus analogous to the grade descriptors typically employed by universities and specifying the quality of students' achievement at progressive grade levels. I see the parallels here with the generic undergraduate grade descriptors of the host business school. These descriptors are not criteria specific to a given piece of assessment; rather they are standards applicable across the board for undergraduate students. They form what Sadler (2005, p.185) would describe as "verbal grade descriptions…with a separate description for each grade level." Akin to 'standards-based assessment rubrics', which are said to provide qualitative descriptions of what is expected of students at different levels (Kite & Phongsavan, 2017), these grade descriptors provide statements of academic achievement standards (levels of quality) at different grades, comprised of expected qualitative properties that characterise selected levels. Such qualities or characteristics as listed must be present for the student's work to reach a particular standard (Sadler, 2014). Criteria are embedded within the descriptions of achievement at each grade (Sadler, 2005). Effectively then, criteria are amalgamated into predetermined fixed standards against which each student's work is judged (Lok *et al*., 2016).

Whilst I acknowledge Sadler's work, and the crossover between his description of 'standards' and the grade descriptors of the host business school, I will be using the term assessment criteria going forward, acknowledging its apparent use as a catch-all term in much of the literature, and that this is how participants in this study referred to the grade descriptors.

Efforts at improving assessment consistency are often directed at increasing the detail and clarity of published criteria (O'Connell et al., 2016). That this should largely solve the problem of inconsistent assessment is implicitly endorsed by the QAA, on the assumption that criteria can be written unambiguously and then applied consistently to non-standard assessments (Tan & Prosser, 2004; Yorke, 2011). This assumption does not hold, which is the fundamental problem with the process. Criteria are of limited power as agents of consistency because their meaning is "fuzzy" (in the words of Sadler, 1989, p.124), and so not effectively or reliably communicated (Bloxham, 2009; Bloxham et al., 2016; Ecclestone, 2001; Price & Rust, 1999; Sadler, 2009). Reconstruction of criteria using language that is more detailed does not necessarily bring fuzzy standards into sharp relief (Sadler, 1987).

Criteria may be codified but this masks the complexity involved. The challenge of assessing against and across several overlapping criteria is severe (Yorke, 2011). Irrespective of how detailed assessment criteria may be, that a degree of professional judgment is exercised is inescapable, if only to resolve the aforementioned fuzziness of the criteria (Bloxham, 2009; Yorke, 2011). Moreover, criteria are interpreted through the lens of the assessor's personal experience by reference to *inter alia* "specialist knowledge, assessment by others, students' work, involvement in moderation and examiners' feedback" (Bloxham & Price, 2015, p.204). Inevitably, understandings, interpretations and the ways in which criteria are combined can and do vary, as do habitual assessment approaches (Sadler 1987, 2005, 2009). Of particular note for this thesis, Webster et al (2000) demonstrated this specifically with respect to assessment of dissertations, highlighting considerable variation and ambiguity in assessors' understandings of criteria as worded.

Irrespective then of how well defined criteria may be, academics may still arrive at different grades for the same piece of work whilst applying the same criteria (Grainger et al., 2008). Variability may also occur because academics hold different conceptions of the criteria; or they may weight and balance criteria components differently; or they may disagree on the use of the criteria and so fail to adopt it; or they may simply ignore it, choosing to apply their own personal standards or tacit or internalised criteria (see, for example, Baume et al., 2004; Bloxham, 2009; Bloxham et al., 2016a; Ecclestone, 2001; Hunter & Docherty, 2011; Price, 2005; Price & Rust, 1999; Orrell, 2008; Read et al., 2005; Shay, 2005; Smith & Coombe, 2006; Webster et al., 2000; Woolf, 2004). Tacit or internalised criteria are a particular barrier to consistency because these are "locked inside the marker's head" (Grainger et al., 2008, p.135). Formed from each academic's predilections, these are not necessarily taken from official documentation, not espoused nor shared with colleagues, and it is not possible to render explicit the tacit knowledge involved in assessor judgements (Baume et al., 2004; Hunter & Docherty, 2011; Orr, 2007; O'Donovan et al., 2008; Sadler, 2009; Sambell & McDowell, 1998). Thus, attempts at externalising knowledge that is essentially tacit and interpretive through delineating evermore-precise definitions of criteria are fruitless (Sadler, 2014).

These issues obviate reliance on assessment criteria, howsoever clearly stated, as *the* route to achieving comparable and consistent judgements (Bloxham et al., 2015; Yorke, 2011). However, several authors promote socialisation processes as solutions, processes that facilitate the diffusion of tacit beliefs and knowledge within a local assessment community (see, for example, O'Donovan et al., 2004; Price, 2005). This brings me to internal moderation.

3.4.2 Internal moderation

Internal moderation processes are fundamental to the mechanisms involved in quality assuring the cycle of teaching, learning and assessment in higher education institutions (Beutel et al., 2017; Klenowski & Wyatt-Smith, 2013). Moderation, in a broad sense, is the comparative practice of two or more academics appraising students' work with the aim of improving consistency in assessment and ultimately assuring that the grade awarded to each student is independent of which academic actually completes it (Grainger et al., 2016; Sadler, 2013).

Moderation operates through academics discussing and debating the quality of assessed work (Klenowski & Wyatt-Smith, 2013). This can involve 'second marking', where an academic assesses a set of students' scripts that are then reviewed on a sample basis by another assessor; or 'double marking', where two assessors each assess the same set of students' scripts before comparing grades. Both approaches involve collective engagement by pairs or teams of academics who may therefore develop shared interpretations of standards and shared understandings of what constitutes evidence of performance at different grades (Grainger et al., 2016; Smith, 2012). Moderation thus forms an important element of socialisation processes through which academics learn of and collectively develop local standards (Reimann et al., 2010).

All that being said, studies have shown that moderation fails to eradicate discrepancies between academics, with inter-assessor correlations observably low where empirically tested (see, for example, Jonsson & Svingby, 2007; Meadows & Billington, 2005; Williams & Kemp, 2019). Additionally, substantial differences in grades awarded in isolation by separate academics for an essay are common (Partington, 1994). Moderation arguably then fails to meet its objectives (Bamber, 2015). Power imbalances

appear to form a key constraint to effective moderation (see, for example, Bloxham, 2009; Bloxham & Boyd, 2007; Orr, 2007; Partington, 1994; Sadler, 2009, 2010, 2013; Reimann et al., 2010). Inevitably, more senior/authoritative/experienced academics hold sway over more junior colleagues who often adopt a safety first approach of simply converging their grades (Partington, 1994). Similarly, in larger groups with multiple assessors inexperienced members have been observed to go with the majority. Is this acculturating to a local assessment community? Or is it toeing the line of "fairly much what everybody else was awarding" (Handley, den Outer & Price, 2013, p.894)? Either way, it seems that for less senior/authoritative/experienced academics their own judgements are of secondary concern (Grainger et al., 2016). I have also said that academics draw on their connoisseurship when assessing students' work. This further serves to exacerbate imbalances in experience, and so power relations, in forming differing individual conceptualisations that inhibit agreement (Ecclestone, 2001; Grainger et al., 2016; Sadler, 2009). This then threatens the equity so needed and called for in academic judgements (Bloxham et al., 2016b).

3.4.3 External examination

In the UK context, professional self-regulation of assessment standards in higher education is provided by a system of external examination. This assurance system is longstanding and well established as a means of validating the credibility of assessment standards (Bloxham et al, 2016b; Silver & Williams, 1996). It is widely admired for the part it plays in defending and assuring standards (Bloxham & Price, 2015). With external examiners held in high esteem, and the system described as "a leading example of best practice around the world" (Universities UK, 2011, p.5), it is unsurprisingly in widespread use in many jurisdictions including the UK (Bloxham & Price, 2015; Bloxham et al., 2015).

External examination focusses directly on the quality and fairness of assessment processes and assessed outputs (Bloxham & Price, 2015; Gaunt, 1999). Exact roles and remits can vary, however ordinarily this involves "inter-institutional peer review of assessment" whereby an academic in one higher education institution acts as an external examiner for another (Bloxham et al., 2015, p.1071). Through independent review of assessment tasks and students' work (usually by sample) external examiners offer institutions a "reassuring presence" (Webster et al., 2000, p.78) and provide a

"'public defence' of their assessment processes and standards" (Bloxham et al., 2016b, p.647).

However, the system of external examining is subject to claims, at least by some authors, of unreliability and ineffectiveness (see, for example, Price, 2005; Silver, Stennett & Williams, 1995). One of the major criticisms levied is that external examination relies upon shared knowledge of standards that are socially situated, co-constructed and calibrated locally (Bloxham et al., 2015; Shay, 2005). It follows that academics from one institution, when charged with reviewing the processes by which the grades have been arrived at in another institution will draw upon their own practices and expertise and hence, will not necessarily possess the same understandings of institutional standards (Beutel et al., 2017). They will instead apply their own interpretation of standards (Bloxham et al., 2011). This is where the importance of assessment criteria comes in, at least in theory to equalise any such differences leading to reliable and consistent approaches (Colley & Silver, 2005). I have previously covered the concerns over reliance on assessment criteria as a panacea for assessment consistency. Here too research has shown that assessment criteria are insufficient in and of themselves to facilitate congruence between internal markers and external examiners (Bloxham et al., 2015). To summarise relevant aspects from prior discussion, this is partly because of deficiencies in criteria as written, partly because assessors interpret criteria differently, and partly because not all assessors actually draw on these in whole or in part. External examiners are not excluded from these factors, with evidence pointing to reliance on personal experience over the documented criteria of the host institution (Colley & Silver, 2005). This is why some claim that external examination gives only an impression of external review, which challenges the notion of and confidence placed in it as integral to UK mechanisms for providing quality assurance of programmes of study (Bloxham et al., 2015; Hannan & Silver, 2006; Medland, 2019).

3.5 Conclusion

That assessment in higher education is inconsistent and unreliable is well observed, researched and reported. Even large-scale studies of standardised tests have shown assessor variability is a concern. Considerably more variation is observed when assessing complex higher-order skills such as critical thinking through, for example, essays and dissertations which, being open ended and reliant upon complex interpretive judgment, resist direct measurement. Notwithstanding the causes and evidence of variation highlighted in the body of this chapter, in reality there is a tendency towards outward agreement between assessors in advance of examination boards. However, assessors may not necessarily agree on grades for the same reasons, with questions over shared interpretation of standards, which are much less likely to develop if the thinking that underpins assessment is individually tacit, unarticulated and unshared.

Subject knowledge is important for critical thinking, as is how knowledge is perceived. Critical thinking involves making judgements and this in turn relies upon a recognition that knowledge is constructed and hence fluid and contestable rather than fixed and absolute. Drawing on the epistemological development models of Perry (1970) and Baxter Magolda (1992) can provide some practical indicators of critical thinking in students' work, for example, in the challenging of extant knowledge, the weighing of different opinions and the determination of a position. These models speak then to what is and is not 'critical' in students' work, reflecting academics' expectations.

Institutional mitigating mechanisms such as criterion referencing, internal moderation and external examination purportedly help to develop understandings and facilitate agreement between assessors thereby reducing variability. The accountability paradigm assumes that students' work is most appropriately assessed with reference to criteria; that this minimises the effects of personal judgement, increasing assessment consistency and reliability. However, the research picture is mixed concerning the effectiveness of this. Internal moderation involves collective engagement by pairs or teams of assessors who can develop shared interpretations and understandings, calibrating judgements and reducing variability over time. However, studies have shown that moderation fails to eradicate discrepancies between assessors, with issues of experience, power and connoisseurship working against consistency. External examination too is subject to claims of unreliability and ineffectiveness.

Critical thinking forms an important component of assessed work, however assessment is done by academics whose individual conceptualisations of critical thinking, and whose individual approaches to assessment, may vary. The systems implemented as ways and means of mitigating assessment variability are subject to pros and cons, their effectiveness arguably limited. Perhaps standardisation is impossible to achieve (Bloxham, 2009; Bloxham et al., 2016a). Perhaps, given the complexity of assessing for critical thinking, we simply have to accept the "subjectivity of judgement" (Clegg & Bryan, 2006, p.224) and live with the inevitable variation. Regardless, what this and the preceding literature review chapter have shown is that there is a clear need for determining how academics conceptualise critical thinking and how they operationalise this through assessment. In the next chapter, I explain how I progressed this research detailing my research philosophy, methodology and methods, and taking into consideration applicable concerns for limitations, ethics and rigour.

4. Research philosophy, methodology and methods

4.1 Introduction

Here, in Chapter 4, I present this study's research philosophy, methodology and methods. I begin by outlining social constructionism as the philosophical perspective that underpins and informs the choice of methodology, i.e. case study, and qualitative approach taken to data collection and analysis via semi-structured interviews and thematic analysis. I then look to establish the integrity of this research as planned and executed with reference to its limitations, discussion of ethical considerations attendant to this research, and consideration of criteria for rigour befitting of qualitative research. I end this chapter with a brief summary and look ahead to the findings in Chapters 5 and 6.

4.2 Social constructionism

Social constructionism holds that the social world is not a fixed, objective, observable reality. Actors through their social practices create it (Cohen, Manion & Morrison, 2017). Knowledge then is not simply out there waiting to be discovered, but is actively constructed and constantly reconstructed by social actors through their everyday interactions (Andrews, 2012; Cunliffe, 2008; Schwandt, 2003). Reality is what people make of it, and people and reality are inseparable. "Ideas, thoughts and actions are thus the result of ongoing processes of interactions and interpretations between human beings" (Lindgren & Packendorff, 2009, p.30). Hence, a social practice, such as assessing for critical thinking in the dissertations of undergraduate students (hereafter referred to as 'the practice') is built from the perceptions, actions and interactions of the social actors involved, and is subject to constant revision as new experiences and meanings are socially reproduced (Burr, 2003).

It follows that the focus of social constructionist research is not on individual cognitive processes but on knowledge constructed through social interactions, with the emphasis on investigating social action and deriving meaning from interpreting participants' accounts (Berger & Luckmann, 1991; Young & Colin, 2004). This study, characteristic of qualitative research within a social constructionist paradigm, is exploratory, experiential and focussed on accessing the socially situated contextualised accounts of human participants with respect to a social practice with which they are engaged (Braun &

Clarke, 2006). The practice is not an objective entity with a reality external to the social actors involved. It is not something that I can observe, measure and document in detached fashion in the positivist tradition (Bryman, 2015). Moreover, my concern is not with what the practice is *per se* (a first-order perspective) but on how participants experience it (a second-order perspective). If I were to adopt a first-order perspective, which is typical of research underpinned by an objectivist stance, I would seek to examine the practice directly and describe it as it is. However, as an investigation of how participants interpret their social world, it is their perspectives rather than mine that orient this research (Yates, 2004). I have sought to study their social world from within not abstracted from it, engaging with social actors in their natural setting, attempting to interpret the practice and make sense of the meanings which they attach to it (Denzin & Lincoln, 2008). Taking a second-order perspective is consistent with social constructionism because I seek here to access participants' subjective lived experiences, to ascertain how they understand their world and how they would describe their reality with respect to the practice (Andrews, 2012; Steedman, 2000).

Social constructionism is a perspective that is said to favour a qualitative research approach (Coolican, 2009). Qualitative research draws on the hermeneutic tradition and is thus concerned with subjectivity, with locating meanings and rich, socially situated and contextualised behavioural insights (Guba & Lincoln, 2005). Qualitative research methods, such as observation and interviewing, enable the researcher to get closer to participants' lived experiences and reality relative to more remote quantitative methods that can produce artificial and sterile findings that bear limited applicability to our everyday lives as human beings (Denzin & Lincoln, 2008).

Applying the thinking of Lindgren & Packendorff (2009), knowledge of the practice under investigation, as socially constructed, is knowledge pertaining to how relevant actors (business school academics in this case) individually and collectively recognise, define, construct and re-construct that practice. Discourse, which I delineate simply as textual and verbal communications, is crucial because knowledge is constructed through discourse. People construct meanings and understand their world through language and communication (Burr, 2003). Operating also on the assumption that socially mediated accounts of reality can be accessed through language (Madill, Jordan & Shirley, 2000) the emphasis in this study is on language-mediated communication and the meanings that are attributed in words. This is evident in the qualitative approach taken to data

collection (via semi-structured interviews) and analysis (via thematic analysis) within the overall case study methodology.

I recognise that my choice of research approach and methods should fit with my research questions (Punch, 2005). In fact, these helped guide me in the process of choosing appropriate data collection and analysis methods. Questions drawn qualitatively, as here, open-ended rather than specific or measurable, seeking to explore human perspectives on a socially constructed practice and concerned with the interpretation of meanings, beg for qualitative methods of data collection and analysis (Denzin & Lincoln, 2008). The semi-structured interview format facilitated discursive exploration of the practice, imposing a degree of structure to maintain relevance but also allowing conversations to move flexibly. Thematic analysis, being independent of epistemology and not aligned with any particular theoretical frameworks, is compatible with a range of research paradigms and questions (Clarke & Braun, 2013, 2017; Maguire & Delahunt, 2017). Applied to interview data, this was appropriately open-ended with structuring and conceptualisation occurring through the analytical phase, typical of qualitative research (Punch, 2005).

4.3 Case study Methodology

Case study has become a mainstay of social science inquiry and, of particular relevance to this thesis, of education research (Hamilton & Corbett-Whittier, 2013; Thomas, 2011a; 2013). Case study is not a research method as such but a “strategy for understanding” (Punch, 2005, p.144), a design frame for inquiry that can take in a variety of methods (Simons, 2009; Thomas, 2013). Stake would agree describing case study as not a methodological choice *per se* but more a choice of what to study by whatever methods are appropriate (Stake, 2005; Starman, 2013; Thomas, 2010). Case study is a valuable means by which we can develop as full an understanding as possible of a case or cases and so gain an understanding of a complex phenomenon and key insights into local practices (Cooper & Morgan, 2008; Punch, 2005; Silverman, 2015).

A single case study approach, as differentiated from multiple or cross-comparison case studies, entails a thorough examination of a particular case in depth and in situ, appreciating its context and complexity as integral to our understanding (Punch, 2005; Stake, 2005; Yin, 2011). The case is “an individual unit (as a person or community)” (Merriam-Webster’s dictionary, 2009; cited by Flyvberg, 2011, p.301). This is further described as “a unit around which there are boundaries” (Merriam, 1998, p.27), a “functioning specific” or “bounded system” (Stake, 2008, pp.119-120), or “a bounded entity” (Yin, 2011, p.6). Such a unit could be a particular person or event, or more commonly an organisation or a location, for example a university or one of its constituent faculties or departments, or a social phenomenon (Bryman, 2015; Farquhar, 2012; Yin, 2011). The researcher observes but does not control the situation (Yin, 2018). Particularisation of the case and understanding it in detail are core concerns (Stake, 2005). The aim is to generate a detailed and contextualised understanding of a case situated within its specific bounded real-world context (George-Williams et al., 2019; Hancock & Algozzine, 2017; Yin, 2018). The case supplies the appropriate context, which is vital for understanding, and blending description and analysis of the case helps in answering the research questions (Cohen et al., 2017; Creswell & Poth, 2018; Yin, 2011). This is typically achieved through ethnography or, pertinent to this thesis, by qualitative interviewing (Bryman & Bell, 2011).

I have taken a case study approach in this thesis in order to develop an in-depth understanding of academics’ conceptualisation of critical thinking and its operationalisation through assessment of undergraduate dissertations in a specific higher education setting. This then is the individual unit of study within its bounded context. Following Yin (2018), as a unit of analysis this is suitably concrete, i.e. a real-life phenomenon. In addition, the sort of complexity here, also following Yin (2018), would point to the value of taking a case study approach. This value is, as here, in examining pertinent questions of a contemporary phenomenon in a practical, real world setting (Farquhar, 2012). In so doing, I aim to produce the sort of “concrete, context-dependent knowledge” that others can learn from and which Flyvberg (2011, p.302) contends a case study should provide.

I would categorise the case as ‘intrinsic’ according to Stake’s (2005) typology, because “the case itself is of primary interest” (Gray, 2013, p.266), the fundamental aim being to gain understandings of and insights into that particular case, the “particularities of a situation”, its “uniqueness” and “complexity” (Bryman & Bell, 2011, pp.60-61). All are arguably key strengths of the case study approach (Lee, Collier & Cullen, 2007). This is akin to Merriam’s ‘particularistic’ case study categorisation, with the focus on a particular event or, as here, a particular phenomenon or practice, and the concern for practical input to “questions, situations, or puzzling occurrences arising from everyday practice” (Merriam, 1998, p.29). I also regard this case as ‘instrumental’ per Yin (2018). Here, I am focussing on a practice and have selected a case to illustrate this, i.e. the case is used as a tool to illuminate the practice (Creswell & Poth, 2018; Thomas, 2013). The aim is not just to understand better the particular case (intrinsic or particularistic) but to examine issues of critical thinking and assessment in higher education through that case (instrumental).

I chose the particular case because, following Stake (2005), I considered it offered an opportunity to learn about the practice under investigation in a setting where I anticipated that learning would be great. In addition, I consider the case may be indicative of broader higher education assessment practices, exemplifying a larger set (Bryman, 2015). I say ‘may’ because the inability to generalise from a case study is a long-standing criticism of the methodology (Flyvberg, 2011). Appreciably the findings, or ‘assertions’ per Stake (2005), are not generalizable in the traditional sense, i.e. statistically. However, this is in any event not the purpose of case study research and does not need to be (Bryman & Bell, 2011; Farquhar, 2012; Thomas, 2011a). Knowledge may transfer from the case context to others even though not strictly generalizable (Flyvberg, 2011). The lessons learned, the themes developed from this case, may transcend to others through the “force of an example” (Flyvberg, 2011, p.305), or more formally through ‘analytical generalisation’ (Lincoln & Guba, 1985; Yin, 2018), or more appropriately, as I am not looking to generalise the findings to theoretical proposals, on a comparative case-to-case basis through ‘naturalistic generalisation’ (Stake, 2005). I will return to the issue of generalisability in more detail in s.4.6.4.

4.4 Data collection

4.4.1 Qualitative research interviews

Kvale (2007, p.1) adroitly put it thus, “If you want to know how people understand their world and their life, why not talk to them?” Interviewing follows logically from seeking to gather participants’ experiences and conceptual understandings of a social phenomenon or practice (Sin, 2010). Interviewing thus plays a central role in qualitative research as an accepted way of accessing human experiences, perceptions, meanings and realities, i.e. the ‘lived worlds’ of participants (Brinkmann & Kvale, 2015; Joffe, 2012).

4.4.2 Semi-structured interviews

Semi-structured interviewing is an especially common format in qualitative social research, being particularly suited to gathering data on participants’ perspectives in their own words, and effective in delving down beneath the surface to make sense of how participants actually experience, understand and behave in relation to the practice under investigation (Matthews & Ross, 2010). The format follows a common interview script, protocol, guide, agenda or schedule (Creswell & Poth, 2018; Kvale, 2007). Whilst such terms appear interchangeable, my own preference is for ‘guide’ which I will employ from herein and by which I mean a document comprised of a limited number of key predetermined questions to cover (Creswell & Poth, 2018). These follow from the study’s research questions and are, as is often the case, derived from a detailed review of the literature (Zucker, 2009). The guide helps structure the interview, concentrating attention where it needs to be, guiding the interview so that the object of study remains in focus and that key aspects are covered (Joffe, 2012; King, Horrocks & Brooks, 2019).

Semi-structured interviewing can also usefully combine with other methods of data collection (Matthews & Ross, 2010). In addition to being interviewed I asked for two actions from participants: firstly, to complete and return a questionnaire comprised of open ended questions in advance of a face-to-face interview; and secondly, at interview to provide two recently assessed dissertations of their choice which for them exhibit contrasting levels of critical thinking. My aims, supported by these mediating artefacts, were to facilitate reflection both before and during the interview thereby encouraging meta-awareness in participants to help elicit insights into their experiences.

The semi-structured interview format engineered a degree of commonality across interviews but remained usefully open to surprise, affording the flexibility to cover common questions or topics in different ways and in different orders and to proceed down differing paths to bring out fresh perceptions, and potentially answers to unintended questions (Matthews & Ross, 2010). I certainly found that I could probe for elaboration, encourage ongoing reflection and tailor questions to each participant's concerns and interests in the course of discussion (Yates, 2004). Through active listening and critically following up on initial responses, I funnelled down to tease out more detail and specifics on points of interest, probing for meanings and understandings (Kvale, 2007). I asked questions in different and overlapping ways around the core issues of critical thinking and assessment covered by the guide, hunting for richer descriptions and further mitigation of my own bias. *Prima facie* this may seem tedious but I would agree with Koole (2012) that asking similar questions in alternative ways and from different angles on and around the core concerns can help to both clarify meaning and elicit different views.

4.4.3 Participant sampling

I determined a need to interview people who would know a great deal about the practice under discussion and who were positioned to offer some insights into that. This meant purposively sampling participants from a bounded population (Farquhar, 2012), i.e. academics in a host business school. Hence, people who were of specific interest, who have experience of the practice, who were likely to possess the necessary information content, and who were best placed to answer the interview questions and by extension the research questions (Bryman, 2015; Creswell & Poth, 2018; Flyvberg, 2011; Joffe, 2012;).

Interview-based studies commonly comprise 15 +/- 10 participants (Kvale, 2007), although there is no universally applicable sample number. I deemed the sample should comprise as many relevant participants as necessary to get the answers needed and that I should continue to interview until reaching a point of saturation, i.e. the point at which "further interviews yield little new knowledge" (Kvale, 2007, p.44). In practice, this is where interviews become repetitious and fresh insights cease to accrue (Malsch & Salterio, 2016), A dozen appropriately selected interviews may be sufficient (Lincoln & Guba, 1985). Beyond this could tend towards redundancy (Dai, Free & Gendron, 2019). I note similarly from interview-based studies in my own professional field of auditing, for

example, that saturation is often claimed to have been reached after twenty interviews, if not before (Malsch & Salterio, 2016).

In my experience, a university business school is a complex social world comprised of several academic disciplines that co-exist under the umbrella of ‘business’ but which exhibit commonalities and discordances running along fiercely independent disciplinary lines. Conscious of this, and of the desire to draw from a pool of people who could potentially offer a spread of relevant experiences, I sought to include participants from across the five disciplines of the host school, namely Accounting, Business and Management, Marketing, Finance and Economics. Ever the pragmatist, I pursued as many as were willing to be interviewed (twenty-one in total). The resultant dataset totalled nearly thirteen hundred minutes of audio-recorded data, averaging some sixty minutes per interview. As the interviews progressed, I did get a sense that later interviews were re-treading previously covered ground, which I took as indicative of having reached saturation from my standpoint as the researcher. This is not to say that the knowledge here is exhausted and further insights would not be forthcoming but I had to draw a line under the process, cognisant of the inevitable trade-off between adding to the richness of participants’ accounts versus the diminishing returns of bearing witness to repeated points.

I conducted all of the interviews in the academic year 2017/18. Relevant background information for each participant is summarised in Table 4.1 overleaf.

Table 4.1 Participants profiles

Participant (Pseudonym)	Academic discipline	Gender (M/F)	Job title	Academic track: Research (R) / Teaching (T)	Research active (Y/N)	Time in University sector (years)	Time in host School (years)
Oliver	Accounting	M	Associate Professor	R	Y	40	3
Fiona	Accounting	F	Professor	R	Y	20	3
Evelyn	Accounting	F	Associate Professor	T	N	24	5
Oscar	Accounting	M	Professor	R	Y	24	5
Hanna	Economics	F	Assistant Professor	R	Y	18	18
Thomas	Economics	M	Professor	R	Y	27	5
Olga	Economics	F	Professor	R	Y	9	4
Harry	Business and Management	M	Associate Professor	R	Y	25	17
Arthur	Marketing	M	Professor	T	Y	19	17
Olivia	Marketing	F	Associate Professor	R	Y	21	4
Isaac	Finance	M	Associate Professor	T	N	19	17
Theodore	Finance	M	Associate Professor	T	N	15	15
Florence	Marketing	F	Assistant Professor	T	N	12	12
Reuben	Business and Management	M	Associate Professor	T	Y	14	14
Mason	Business and Management	M	Professor	R	Y	31	13
Milo	Finance	M	Assistant Professor	T	Y	10	5
Zhao	Marketing	M	Associate Professor	R	Y	20	4
Viola	Economics	F	Assistant Professor	R	Y	19	17
Ella	Marketing	F	Assistant Professor	R	Y	14	5

Participant (Pseudonym)	Academic discipline	Gender (M/F)	Job title	Academic track: Research (R) / Teaching (T)	Research active (Y/N)	Time in University sector (years)	Time in host School (years)
Aidan	Accounting	M	Assistant Professor	R	Y	16	8
Penelope	Business and Management	F	Associate Professor	T	Y	11	11

4.4.4 Pilot interviews

I conducted two pilot interviews with current academics. The first was situated in Durham University School of Education (Pilot 1) and the second in the host business school (Pilot 2). Having gained their preliminary verbal acceptance, I subsequently sent pilot participants a Participant Information Sheet (Appendix B). This explained the aims of the study, the interview process and covered information necessary to obtain informed consent. Completion of a pre-interview questionnaire preceded each pilot interview.

For Pilot 1 the pre-interview questionnaire (Appendix C) comprised four open questions focussed on what critical thinking is, how critical thinking is evidenced in students' work and how this had come to be known. I reviewed the return shortly prior to interview, annotating with comments and follow-up questions that then informed the discussion. I further subsumed the pre-interview questionnaire within the interview guide. This covered three main areas: firstly, the assessment process; secondly, critical thinking; and thirdly, a discussion of assessed pieces of students' work.

For Pilot 2 I subsequently expanded the pre-interview questionnaire (Appendix D) to five questions covering the same areas as before but with added emphasis on how understanding of critical thinking had developed over time, which was something interesting that came out of Pilot 1. This change was accommodated in the guide which otherwise remained unchanged save that the reference to 'undergraduate students' work' was amended to 'undergraduate dissertations'. I always intended undergraduate dissertations to be in focus for this study. That Pilot 1 centred upon 'undergraduate students' work' more broadly was simply a practical imperative with Pilot 1 engaged in assessing assignments not dissertations. This notwithstanding, I derived much value from the pilot stage as a dry run of the interview process.

Conducting pilot interviews enabled me to check that I would be communicating with sufficient clarity and friendliness of tone and that data collection would cover relevant areas. Feedback and review served to establish the trustworthiness of the pre-interview questionnaire and interview guide. In particular, following Yin (2018), I considered the focus and nature of the questions I was asking to determine if the process was likely to elicit relevant responses and to refine the interview guide accordingly. Notably, the two pilot participants provided apparently disparate levels of response to the pre-interview questionnaire. Pilot 1 offered quite considered and detailed responses to the original four set questions, whereas Pilot 2, who described the enlarged five-question format as cumbersome and repetitive, provided only minimal responses. I therefore reverted to four questions. I also thought, with the benefit of hindsight, that some of the interview questions, whilst comprehensible, seemed lengthy and somewhat formal, noting that I had abridged these on an ad-hoc basis during the course of the interviews. I was at least comforted that the questions were scheduled in a logical order as intended to help maintain the structure and flow of the interview in moving from point to point (Saunders, Lewis & Thornhill, 2012; Yates, 2004).

4.4.5 Main interviews

I made initial targeted requests for participation across the host business school. Where preliminary verbal assent to participate was received, I followed this up with an email invitation that included a copy of the Participant Information Sheet and pre-interview questionnaire. I subsequently arranged interviews at each participant's convenience. In all I might have had up to thirty participants, however scheduling interviews proved difficult in the case of several initially willing respondents. I made a decision not to persist beyond two further polite email requests. I issued each of the twenty-one participants in the main interviewing phase with a copy of the pre-interview questionnaire. Seventeen participants completed and returned this.

I took appropriate advantage of the flexibility offered by the semi-structured interview format so that whilst each interview started from the same point and loosely followed the interview guide (Appendix E), I did not feel constrained by that and did not allow it to inhibit the conversation. I let interviews flow in whatever direction the participant chose to pursue whilst at appropriate junctures reverting to the guide to keep things on track.

Consequently, no two interviews were structurally identical. However, I did seek to cover all areas of the guide in the course of each interview.

I structured the guide along three progressive sections. The first served a number of important purposes. It anchored the interview firmly in a situated context and oriented the early part of the discussion around assessment in higher education and more specifically of undergraduate dissertations. The first question also served as a relatively gentle introduction to the interview, offering participants an open invitation to talk about their personal background, which is something they would obviously know about (Matthews & Ross, 2010). This helped with rapport building and with putting participants sufficiently at ease to engage in a hopefully honest and open discussion. The second section reprised the pre-interview questionnaire. Here, the interview turned towards critical thinking, revisiting the areas covered by the questionnaire, seeking to draw out how participants would define and characterise critical thinking both generally and with respect to undergraduate dissertations. The third section built on the first two, turning attention to assessing for critical thinking in undergraduate dissertations, placing the focus firmly on the practice under investigation. This was facilitated by document elicitation (refer to s.4.4.6 for more detail). I concluded each interview with brief thanks for taking part and reiterated assurances of confidentiality (Matthews & Ross, 2010).

4.4.6 Document elicitation

Document elicitation draws on photo elicitation as an interview-based method for gaining insights into participants' conceptions and approaches (Buswell & Berdanier, 2020). Photo elicitation developed in the disciplines of anthropology and sociology. Harper (2002) provides a detailed account of this and explains how he used the technique to good effect in his research by showing photographs to interviewees and asking them questions whilst doing so in order to evoke feelings, memories and responses that may not otherwise develop from discussion alone. Similarly, I sought to use dissertations as documents which participants could refer to and talk me through, thereby delving deeper into critical thinking as concept and as assessed in practice.

Wherever possible I timed interviews to coincide with undergraduate dissertation assessment in the academic calendar. Beforehand I asked participants to identify two examples of recently assessed undergraduate dissertations, one of which for them rates highly for critical thinking and one of which rates less so. At the document elicitation stage of the interview, I would say to the participant along the lines of, “For each of the two dissertations you have identified, can you describe the differing levels of critical thinking, how this is evidenced for you and how this has influenced your grade?” In the main, I allowed the participant to speak at this stage, only interjecting where the conversation lulled or where further probing was needed as the conversation developed.

Buswell (2018) in her study found document elicitation to be a very useful way of drawing out rich information of the teaching conceptions held and methods employed by assistant engineering professors. Indeed, she claimed that the elicitation stage of each interview was comparatively more revealing than other stages, professing it “extremely successful in capturing a “thick description”” (p.6). This interviewing technique shows promise and I anticipated then that this would prove a fruitful part of each interview. However, this did not quite turn out to be the case, with the technique not working as hoped for this study. In the two findings chapters I have identified utterances from the document elicitation stages of the interviews via the use of a description in square brackets **[…]**. These utterances are often enlightening but disappointingly few. In part this was due to not all participants having assessed dissertations to hand for the interview, my attempts at timing notwithstanding. However, I submit the greater fault lies with timing in interviews. Situating the document elicitation stage at the end of each applicable interview meant that it became something of a victim of interview fatigue. Consequently, participants gave this stage less attention than they might have done if I had foregrounded document elicitation earlier in the interview guide and made it more a focal point. Still, I feel some insights did come of this and it was worth doing.

4.5 Data analysis

4.5.1 Thematic analysis

Thematic analysis is an analytical method that provides a transparent and systematic approach to interpreting qualitative data (Joffe, 2012; Nowell, Norris, White & Moules, 2017). Put simply, it offers a way of meaningfully processing raw qualitative research data (Matthews & Ross, 2010), comprised in this study of semi-structured interviews. Processing involves organising the data through coding and making interpretations, i.e. "making sense of the data" (Creswell & Poth, 2018, p.195), with the prospect of identifying, analysing and interpreting patterns leading to the construction of themes (Braun & Clarke, 2006).

The approach might be classed as inductive, with themes driven by what is in the data unencumbered by "extant theoretical frameworks" (Vaismoradi, Jones, Turunen & Snelgrove, 2016, p.106) or any "analytical preconceptions" or a "pre-existing coding frame" (Braun & Clarke, 2006, p.12). The researcher thus derives meaning from the data from the 'bottom-up' as opposed to a 'top-down' application of existing theory (Braun & Clarke, 2012). The analysis is then grounded in the data and the researcher frequently returns to the raw data to check the interpretations and to reconsider linkages formed (Matthews & Ross, 2010). However, the development of themes is in reality likely to be informed by theory, as research is not conducted within a vacuum. The researcher will have done some reading, the extent of which can vary, but this will inevitably inform their interpretation of the data leading to themes (Howitt & Cramer, 2014).

In practical terms the approach falls somewhere between 'bottom-up' and 'top-down', with coding and subsequent analysis combining the two to a greater or lesser extent subject to the epistemological orientation and exact approach of the individual researcher (Braun & Clarke, 2012). Certainly, whilst I strived to analyse my dataset inductively for themes, my analysis was inevitably informed by extensive prior reading of the literature and its themes. I actually see this as an advantage. Not only did this help sensitise me to subtleties within the data (Tuckett, 2005), it enabled me to appropriately devise research questions for the study and adequately make sense of the data with respect to those questions. These things follow from understanding the literature and grounding the study in that understanding. Appreciably, however, this needs to be balanced with allowing participants' accounts to speak, taking suitable care to enable

that without forcing a literature-led interpretation. For me, the benefits outweigh the potential downsides of narrowing my analytical scope to that which I might anticipate from my reading of the literature, and which would potentially inhibit my ability to develop themes (Vaismoradi et al., 2016). In effect, immersion in the literature allowed me to engage in a more learned and critical analysis, building on what might otherwise have stalled at the simplistic and the observational.

Braun and Clarke (2006) offer a robust, sophisticated, systematic and comprehensive framework for thematically analysing qualitative data in relation to research questions. This framework involves six phases: (1) familiarising yourself with the data; (2) generating initial codes; (3) searching for themes; (4) reviewing themes; (5) defining and naming themes; and (6) producing the report. These phases proceed linearly but it is recognised that I as the researcher will revert back to earlier phases repeatedly and throughout to clarify and refine themes (Howitt and Cramer, 2014; Vaismoradi et al., 2016).

4.5.2 Phase 1: Familiarising yourself with the data

In Phase 1, I started the process of getting to know my data, gaining initial familiarity through reviewing pre-interview questionnaires and then personally conducting each interview. This also followed from conducting a relatively detailed if incomplete literature review. I came then to the task of analysing the data with a degree of prior knowledge and some preliminary thoughts (Braun & Clarke, 2006). Some formative pattern recognition did occur as interviews progressed and each new one built on the former and I made notes of my ideas as they formed. These did not prejudice full analysis of the data subsequently but served as useful starting points for data coding and later comparison as I developed the coding layers through several iterative stages.

4.5.2.1 Transcribing interviews

Interviews were audio-recorded and transcribed into text form for the purposes of analysis, as is common in qualitative research (Kvale, 2007). This was a fundamental part of the process of my becoming familiar with the data (Bird, 2005; Howitt & Cramer, 2014). To be exact, I conducted all of the interviews and transcribed the pilot interviews and the first eight main interviews in full. A transcription provider initially transcribed the thirteen remaining interviews. I then checked these in full for accuracy against the original recordings, amending extensively as required. This occupied a lengthy period and was at times tedious and frustrating in equal measure, but served to inform my understanding of the data benefiting the early analytical stages.

Following the example set by Braun and Clarke (2012) I read the full set of transcripts twice, and following Clarke and Braun (2013, 2014) I listened again to each interview, making jottings in the margins and notes of any initial observations. This active process of repeat reading and listening helped me to become sufficiently immersed and ready for the next phase of analysis (Maguire & Delahunt, 2017; Nowell et al., 2017).

4.5.3 Phase 2: generating initial codes

In Phase 2, analysis begins through data coding (Braun & Clarke, 2012). Initial coding involves the researcher working through transcriptions systematically, interview by interview, line by line, scrutinising what the data is saying and summarising key aspects with "pithy label(s) that capture something interesting about the data" (Clarke & Braun, 2014, p.6627). At this stage, the intention is not to capture themes but simply to highlight sections of text that are of interest, labelling and indexing any features that are potentially pertinent (Braun & Clarke, 2006). Highlighting ideas usefully goes beyond counting words or phrases in the nature of content analysis (Namey, Guest, Thairu & Johnson, 2008). Data is thus aggregated and clustered into significant groupings enabling connections to be made across the data (Tuckett, 2005; Vaismoradi, Turunen & Bondas, 2013).

My initial attempt at coding was open and wide ranging and involved tagging data portions with simple descriptive codes of single words or short phrases, with only limited interpretation. This generated a list of fifty-eight initial codes (Coding 1st round – see Appendix F).

4.5.4 Phase 3: searching for themes

In Phase 3, analysis progresses from codes to themes (Braun & Clarke, 2012). This involves sorting and aggregating the long list of codes, inclusive of all relevant data excerpts, into a limited number of common themes (Creswell & Poth, 2018). Themes are "patterns of meaning" (Braun & Clarke, 2012, p.57) and "key characters in the story we are telling about the data" (Clarke & Braun, 2018, p.108). Themes should have something important to say about the data, providing insights in relation to a study's research questions (Spencer et al., 2003; Vaismoradi et al., 2016).

Meaning is developed from analysing across the dataset, with themes drawing from shared content, experiences and patterns in the data (Braun & Clarke, 2006). Analysis is, however, an active process of discovery. Themes are not passively revealed. They do not exist in the data simply waiting to be found. Nor do they just emerge from analysis (Packer, 2017). Themes are actively synthesised from the data by the researcher, grown organically through processes of coding and thematic development (Braun & Clarke, 2006).

In practical terms, my initial coding structure was repeatedly revisited and revised. Initial codes were combined. Areas of difference, similarity, overlap, consensus and conflict were considered. Relationships within and between codes were identified and major patterns were formed, building to higher levels of coding which offer greater coherence and meaning than the initial descriptive codes represent (Fereday & Muir-Cochrane, 2006; Nowell et al., 2017). There was a degree of intuition at play as I devised themes, however, this is tempered through layers of interpretation, "a coding of codings" so to speak (Howitt & Cramer, 2014, p.383).

Starting with the initial list of fifty-eight codes, I expanded, reworked and restructured this formulating ninety codes under four main headings and twelve sub-headings (Coding 2nd round – see Appendix G). Some codes were discarded, new codes were formed, superordinate and subordinate headings accrued as analytical ideas developed and relationships in the data and nuances in meanings became more apparent. I then reviewed this second round of coding at length, reading collected codes in isolation but also moving back and forth to the original data for context and understanding (Braun &

Clark 2006). This next round was a refinement of the previous, more with the tagging of interviews / data excerpts than with the coding structure which remained much the same and therefore still overly long and complex (Coding 3rd round – see Appendix H). Agreeing with Howitt & Cramer, 2014, I saw these early rounds very much as a process of trial and error, finding them increasingly over-coded, thereby obscuring what the data really had to say and complicating rather than simplifying the picture. I worked then on decluttering and simplifying, collapsing and amalgamating codes to form a much reduced structure comprised of five main headings, constituting early themes, with between seven and nine codes attached to each heading totalling forty codes (Coding 4th round – see Appendix I) with near nine hundred data excerpts tagged to those codes.

4.5.5 Phase 4: reviewing themes

Phase 4 is a review phase. Here, themes are reconsidered with reference to the data as coded and the dataset in its entirety. This is effectively a quality control mechanism. It is a self-check on the themes as I have interpreted them from the data (Clarke & Braun, 2013).

At this point my tentative themes needed refining and re-examining against the data to ensure each held a coherent pattern (Nowell et al., 2017). I scrutinised them in detail, judging for evidence, i.e. examining whether the data really supported the themes, as well as coherence and fit, both at the level of each theme's constituent coded data excerpts and with respect to the dataset as a whole (Maguire & Delahunt, 2017).

This phase spanned several repeat reviews as I sought to delineate what makes each theme distinctive and how the themes operate together to form a coherent whole. I developed the results of this into a further round of coding (Coding 5th round - see Appendix J). This process of review, reflection and re-working was invaluable, helping me to focus on the salient meanings, and in seeing and judging my early efforts in developing themes as precipitous and in need of further thinking. As a result, I gained in confidence that I have not forced a particular coherence upon the data and that the story I ultimately tell here has developed justifiably from it.

4.5.6 Phase 5: defining and naming themes

In Phase 5, I was aided in further reviewing, refining and clarifying the themes through the thought put into defining and naming them. Defining here means detailing the analysis of each theme, identifying its essence, explaining what each theme is about and how the themes fit together, in short form within a concise and informative name (Clarke & Braun, 2014).

A key aspect here was ensuring that themes are distinct, that each is well focussed, that each can stand alone, and that each can be differentiated in terms of what it is and is not through allocated definitions and names (Clarke & Braun, 20144). Again, it was important for me to revert to the original data and test revised themes for efficacy and fit and to corroborate the findings (Braun & Clarke, 2012; Fereday & Muir-Cochrane, 2006).

Careful thought and much revision was given to naming themes, conscious of the need for brevity yet also the need to give the reader a clear and immediate sense of what each theme is about (Braun & Clarke, 2006).

In summary, I synthesised two distinct themes:

- Theme 1: Facets of an undergraduate student as critical thinker (reported in Chapter 5);
- Theme 2: Approaches to assessing undergraduate dissertations: convergence, differences and mitigation (reported in Chapter 6).

4.5.7 Phase 6: producing the report

Phase 6, the writing up stage, accounts for the results of the prior phases leading to defining each theme and determining how the themes interrelate. This account draws on collated data excerpts for each theme to construct a detailed, coherent, analytical and explanatory narrative, individually and collectively, i.e. the story that each theme tells situated within the overall story (Braun & Clarke, 2006). Themes collectively narrate the story of the data as interpreted and as applied to the research questions (Clarke & Braun, 2014). Hence, themes should clearly relate back to the research questions posed at the outset of the study (Howitt & Cramer, 2014). This is important because the process does not entail analysis for analysis sake. It must have a purpose, which is to answer research questions through finding and examining patterns of meaning across the dataset that are relevant to those questions (Vaismoradi et al., 2016).

Data excerpts should appropriately support the analysis (Braun & Clarke, 2012). In Chapters 5 and 6, I have woven interview quotations into the analytical narrative, sufficient to illustrate the themes and confirm the links between my interpretation and the evidence (Clarke & Braun, 2013; Greenhalgh & Taylor, 1997). I attribute quotations to individual participants, each by their allocated pseudonym, and presented in italics. On occasion my own input as interviewer is included in square brackets […] simply to situate quotations within interview dialogue, for example, where a participant's response makes indirect references which require prefacing to aid the reader's understanding.

In Chapter 7, I have further refined my analysis through consideration of the themes in the context of the extant literature base and this study's research questions. Engagement here was fundamental to me challenging my own interpretations in light of potential alternatives and in developing the meanings and implications of the themes that I have synthesised from the data (Marshall & Rossman, 2016).

4.6 Limitations

4.6.1 Interviewing

Interviewing is, as stated, often the data collection method of choice for the qualitative researcher (Rands & Gansemer-Topf, 2016). However, reliance on interviews is not without criticism. For example, the weight placed on interview data is of particular concern (Saljo, 1997). Saljo (1997, p.176) refers to the failure of researchers to recognise the "primacy of talk", i.e. what interviewees say and what they actually think may not necessarily accord. It is possible, for example, that what interviewees say is what they think the interviewer wants to hear, or expects to hear, or is an attempt to preserve face rather than give an honest answer (Saljo, 1997). Yin (2011, p.13) further refers to participants potentially "echoing the same institutional "mantra", developed over time for speaking with outsiders" rather than giving honest accounts of their practices relative to the organisation's practices. Interview data should be approached with caution then as utterances may simply embody ways of speaking rather than be indicative of actual experiences.

Interviewing is essentially a social interaction through conversation necessitating direct communication and interaction between researcher and participant (Rubin & Rubin, 2012; Warren & Karner, 2015). Interviews by their very nature often exhibit asymmetric power between interviewer and interviewee (Saljo, 1997). The fear is that the interviewer may overly influence the discussion, leading participants rather than allowing them to impart their experiences unfettered. This is a weakness of the method which I, following the example of Sin (2010), looked to counter by listening empathetically and attentively; by giving interviewees time and space to respond to questions, resisting where possible the urge to leap in and fill pauses too soon; by verbalising any assumptions I made during the discussion and asking for clarification with follow-up questions; and by trying to avoid asking leading questions. My intention was to focus on the practice as experienced, to give participants space to explore detail and nuances, to resist leading or conditioning their responses.

I further sought to cover issues in repeat fashion and from different angles where possible, through reprising main questions from the pre-interview questionnaire and through probing. Comparisons could then be made of what interviewees said in relation to particular issues. In addition, I strived to build rapport and maintain an attentive,

interested and respectful atmosphere conducive to openness (Kvale, 2007). My efforts were directed at overcoming the primacy of talk and achieving good correspondence between researcher and participants in the account, adding to the credibility of the research (Bryman, 2015).

4.6.2 Transcription

With all due respect to Kvale's concerns as to the reliability of written transcriptions, as decontextualized from oral interviews, and Barnacle's concerns as to communicative reduction from the translation, how reliable transcription is really depends upon its purpose, i.e. the uses to which the transcribed interviews are to be put. My analytical focus is on participants' experiences and their meanings as conceptualised and understood by them, hence, verbatim transcription is appropriate, acceptable and perfectly suited to the task of thematic analysis (Braun & Clarke, 2006). More complex transcription systems, for example Jefferson, which provide much greater transcribed detail, are not necessary for this (Howitt & Cramer, 2014). The additional information provided is simply not required to answer the research questions of this thesis, hence, the extra effort would be wasted and is therefore to be avoided. My focus is on what was said not how. It is unnecessary therefore to construct transcriptions with pauses, disfluencies, tonal inflections, body language, etc. What was important is that the words as spoken were accurately captured in transcription because this formed the basis of analysis.

Analysing transcriptions is fraught with challenges too. Transcription converts verbal discourse into written text. This is itself an interpretive act. Kvale (1996) describes transcription as an act of translation through which interviews are decontextualized leading to the loss of essential meaning. There are structural and operational differences of speech and written text according to their respective communicative rules and conventions. Kvale (1996) therefore warns that meanings are contextual and may not adequately translate with the change of medium. Meanings may also be lost because aspects of experience may not or cannot come across from the written text compared with face-to-face interaction, which offers a much richer communicative event (Barnacle, 2005). Dortins (2002) argues that meaning is not just lost, rather the researcher actively re-contextualises the interview effectively opening up different interpretations. Thus, in relying solely on transcribed interview data I may misinterpret participants' experiences (Sin, 2010).

Following Sin (2010), I took steps to mitigate this by reflecting on each interview shortly afterwards, making appropriate contextual notes, and by personally engaging with the transcription having listened to each audio recording several times. Repeat listening also formed part of the data analysis process as a means by which contextual features could be brought in to facilitate and check for appropriate interpretations and understandings. I also produced updated transcripts of each interview having redacted personal identifiers and any information that could potentially be used to identify participants (Matthews & Ross, 2010).

4.6.3 Sampling

As outlined in s.4.4.3, I sought to construct a sample of participants from people who would know a great deal about what was being investigated and who were suitably positioned to offer some insights into that. This meant purposively sampling participants from a bounded population, i.e. academics at the host business school. From initial targeted requests for participation across the school I might have had up to thirty interviews, although for various reasons this settled on twenty-one actual interviewees. This is still a substantial number. Moreover, during the course of conducting the interviews I felt I had reached saturation, i.e. the point at which “further interviews yield little new knowledge” (Kvale, 2007, p.44). One nagging doubt remained, however. The participants were people whom I might have anticipated would agree to take part. They came from what I would regard as the circle of the willing, i.e. academics, whether research and/or teaching focussed, who are generally positive about students and the student experience, who were interested in my research and who were likely to be prepared to get involved. There are others at the host business school perhaps whose reputations as interesting, eccentric, even difficult characters precedes them. Their voices, their perspectives would no doubt add to the richness and colourfulness of the dataset. Unfortunately, despite targeted efforts I could not garner any participatory interest in those individuals.

4.6.4 Triangulation

The case study method necessitates the collection of sufficient data to enable the exploration of major aspects of the case and to support interpretations offered by the researcher (Bassey, 1999). Hancock and Algozzine (2017, p.16) describe case studies as “grounded in deep and varied sources of information”. Yin (2011, 2018) talks of the likely need for multiple sources of data and data collection methods forming triangulated evidence, for example interviews combined with observational data, without which the trustworthiness of the data may be open to question, or at least difficult to test for. Triangulation would strengthen confidence in the data as interpreted by the researcher (Bassey, 1999; Creswell, 2018; Farquhar, 2012).

My approach did not involve the triangulation of different data sources to the extent advocated by Yin (2018). Moreover, from my epistemological standpoint I do not accept that convergence of evidence though triangulation necessarily leads to ‘facts’ or the ‘truth’. I acknowledge, however, that a degree of triangulation can potentially add to the richness of the picture portrayed by the research and therefore my understanding of it (Remenyi, 2013). To that end, I did bring certain institutional documents into my interpretations of the data collected through semi-structured interviews. I also attended and observed an annual dissertation supervisors’ communication meeting. I also employed mediating artefacts – a pre-interview questionnaire and recently assessed dissertations - to facilitate reflection both before and during interviews, encouraging meta-awareness in participants to elicit insights into their experiences. I did consider asking participants to ‘talk aloud’ whilst assessing dissertations, also asking for permission to observe meetings between first and second assessors to discuss dissertation grades, so that I could observe assessment processes in play. However, operational limitations notwithstanding, I felt this would have impinged on participants’ goodwill, potentially reducing the sample. On balance, I determined that incorporating a discussion of recently assessed dissertations into interviews would produce relevant data for my purposes. Ultimately, I feel that twenty-one interviews yielding almost thirteen hundred minutes of transcribed data, comprised of different voices on the same or similar issues and hence offering a degree of ‘informant triangulation’ (Remenyi, 2013), provided a sufficiently rich dataset for analysis and interpretation. I submit that collectively the approach has supplied information and evidence enough to facilitate my understanding of the case relative to the research questions.

4.6.5 Generalisability

The value of a piece of research, its usefulness in my estimation, can be measured in terms of its generalisability, i.e. the extent to which its findings are applicable beyond the study at hand more broadly to other contexts, situations, etc. (Cohen et al., 2017). Meanings are not context free and cannot be extrapolated beyond the context of a case study (Schwandt, 2003). However, “knowledge may be transferable even where it is not formally generalizable” (Flyvberg, 2011, p.305).

In this thesis, I have taken a case study approach in seeking to develop an understanding of the conceptualisation of critical thinking and its assessment in undergraduate dissertations by business school academics. The value of the approach is here in examining pertinent questions of a contemporary phenomenon in a practical, real world setting (Farquhar, 2012). Appreciably my findings, what Stake (1995) would regard as merely ‘assertions’, are not generalizable in the traditional statistical sense, i.e. “in the usual sense of nomic generalisation, based upon data representative of some population” (Lincoln & Guba, 1985, p.120). However, this is not my intention. What I can aim for is a degree of ‘transferability’ (s.4.8.3 refers). The themes developed and the lessons learned from this case might transcend to others through, for example, analytical generalisation (Lincoln & Guba, 1985; Yin, 2018). Alternatively, and more appropriately here, through naturalistic generalisation (Stake, 2005), by which readers can examine similarities and differences with their own circumstances, judging whether and to what extent the findings from one study may serve as a guide to what might happen in other situations (Kvale, 2007; Thomas, 2013). The findings may resonate then with readers’ experiences, their ‘phronesis’ (Cooper & Morgan, 2008; Thomas, 2011b), helping them to make connections with their contexts and build their own situated understandings (Zucker, 2009).

My findings are localised to a single case setting. I have examined the matter under investigation in a particular institutional context and portrayed the ways in which the participants experience this as I have interpreted them. So whilst my findings may have little in the way of general applicability in the traditional sense, they can shed some light on the assessment of undergraduate dissertations and the conceptualisation of critical thinking and the role this plays therein, albeit in a single case context. Conceivably, the findings could speak for what may be happening in related settings if they resonate with

others (Hamilton & Corbett-Whittier, 2013; Robson & McCartan, 2016). I of course make no specific claims to transferability. Rather, I have provided suitable contextual information of the case, and of how the fieldwork and analysis proceeded, so that readers can judge transferability for themselves. I hope that readers, through engaging in their own reflection on the findings of this case study relative to their own circumstances can make parallels with their own contexts. In this way, this case study can contribute to understanding at a broader level, allowing others to learn from it and potentially informing decision-making, policy and practice (Simons, 1996, 2009).

4.7 Ethical considerations

That the nature of this investigation and the dissemination of its findings could potentially cause harm to participants was something that I took into consideration. Conceivably, participants could suffer reputational harm if my findings were to paint them, their practices or their professionalism in a negative light. The host business school, and the host university by extension, could suffer reputational harm. Broader still this could bring higher education into disrepute leading to a loss of faith in the sector and its assessment practices. I considered this when portraying the findings.

For degree purposes, I appropriately obtained consent from Durham University School of Education's Research Ethics Committee. I then sought organisational gatekeeper consent from the Dean of the host business school before approaching prospective participants. I issued participants with a detailed Participant Information Sheet, which provided relevant information about the research to enable informed choice over whether or not to participate. This information was revisited at the commencement of each interview to ensure consent was given. Repeat emphasis was given to participation being voluntary, that this could be withdrawn at any stage, that interview questions could be declined and that the interview could be terminated at the option of the participant (Kvale, 2007; Yates, 2004). I kept consent under constant review, watching in interview for any signs that might indicate discomfort or distress. Thankfully, I observed no such indicators.

Attending to confidentiality, together with the steps taken to obtain informed consent, would give participants sufficient confidence to take part. I recognise that protecting participants' confidentiality is clearly important (Cohen et al., 2017; Saunders et al., 2012). I made two transcriptions of each interview. The first represented a full and accurate verbatim transcription. The second I adjusted in order to preserve participants' confidentiality. Adjustments included redacting any references to employers or places of work, specific academic departments, colleagues, modules, etc. – essentially any attributive detail that could serve as identifiers. I allocated each participant an anonymous code for storage purposes. Audi-recorded interview data files (mp4 format) were held electronically on a secure server, with access password-protected. I tagged interview transcripts to anonymous codes to preserve confidentiality, with the list of codes filed separately from the participant list. I allocated each participant a pseudonym for reporting purposes, following a well-established research tradition. I stored all data confidentially and in compliance with both Durham University regulations and UK data protection law.

4.8 Research rigour

4.8.1 Trustworthiness

Rigour is a prerequisite for any piece of credible research. Traditionally characterised by validity and reliability, a study that exhibits both will have findings that are reflective of the object under investigation (Sin, 2010). Internal validity (how sound the conclusions generated are), external validity (how generalizable the findings are) and reliability (how accurately the study and its findings can be replicated) are some of the key criteria by which the quality of much research, in particular quantitative research, can be assessed (Bryman, 2015). Qualitative researchers, however, come from a greater mix of ontological, epistemological and methodological traditions (Guba & Lincoln, 1994). Social constructionism, underpinning this thesis, does not have as its purpose the discovery of objective facts or verifiable truths (Burr, 2003). All accounts of social actors are situated socially, historically and culturally. All knowledge derived from social constructionist research is temporary and open to interpretative challenge (Burr, 2003). This means that for my research positivist concepts of validity and reliability are not appropriate for judging rigour.

This of course does not give me an excuse to abandon rigour. Alternative ways of achieving rigour in research do exist. ‘Trustworthiness’ is concerned with the integrity of qualitative research (Guba & Lincoln, 1994; Lincoln & Guba, 1985) and may be split into aspects which match traditional criteria for research rigour. These include ‘credibility’ (are the findings believable?) mirroring internal validity; ‘fittingness’ or ‘transferability’ (can the findings be applied to other settings?) mirroring external validity; ‘auditability’ or ‘dependability’ (can the findings be replicated?) mirroring reliability; and ‘confirmability’ (has the researcher sought to mitigate the impact of his biases and values?) mirroring objectivity (Bryman, 2015). These are pragmatic choices that enable the researcher to demonstrate the rigorous way in which qualitative research has been undertaken to allay concerns over the acceptability of the findings (Nowell et al., 2017).

4.8.2 Credibility

Establishing credibility involves conducting research in accordance with good practice, demonstrating consistency between object, data and conclusions (Sin, 2010).

Context has an important role to play in addressing credibility concerns. Qualitative research is situated and people are often studied in their natural settings (Denzin & Lincoln, 2011; Punch, 2005). Early studies in the phenomenographic tradition, for example, put participants firmly in situations which contextually were meaningful to them and which allowed interviewer and interviewee to jointly share in the experiences of the latter. The results become meaningful when the interviewee is placed in a situation of looking to achieve something within their own context, rather than being targeted with questions in abstract (Saljo, 1997). In this study, I interviewed academics situated in their own business school setting and adjacent in timing to the process of assessing dissertations.

To help maintain interview context both the original audio recordings and the transcribed interviews were uploaded to QSR NVivo, a qualitative coding and analysis software package. This was so that I could move repeatedly between the two, comparing and checking for understandings. Using NVivo helps in managing the data by enabling the tagging and collation of data excerpts through coding. This made searching and sorting the data easier, helped by organising through structuring into groups of similar subjects and ideas and encouraging active reading of the data, thereby aiding comparisons and pattern recognition and hence the depth and intricacy of the analysis (Joffe, 2012;

Nowell et al., 2017). NVivo also provides an audit trail covering the stages of analysis. This evidences rigour, helping to establish the aforementioned trustworthiness of the findings. Note, however, that NVivo is simply a tool for organising, structuring and reviewing a relatively large dataset. I as the researcher, not the software, do the actual coding and categorising (Creswell & Poth, 2018). Nvivo would make no judgements and nor would it transform the data through to themes (King, 2004). It simply helped me to perform these vital tasks and to an extent freed me to do "the hard analytical thinking" (Marshall & Rossman, 2016, p.228). The only downside I found, agreeing with Creswell & Poth (2018), was the time investment required to get up to speed with NVivo, which I did via an introductory workshop, which I attended twice, and working through several guidance books. Even so, this was limited because I concentrated on operating only the basic functions that served the necessary purposes. I did not engage with functionality of the software at a more sophisticated level than I needed. A further argued downside is that of putting some distance between researcher and data, with the data mediated through a software program (Bazeley & Jackson, 2013). I sought to mitigate this through prior reviewing and annotating of transcripts.

4.8.3 Transferability

Transferability in qualitative research surrogates for generalisability in quantitative (Nowell et al., 2017) – "knowledge may be transferable even where it is not formally generalizable" (Flyvberg, 2011, p.305).

A small-scale qualitative study by its very nature is particularly prone to the criticism of lacking much if any generalisability (Bryman, 2015; Flyvberg, 2006, 2011; Thomas, 2010, 2011b; Yin, 2011). However, qualitative research findings are not meant to be statistically generalizable and no such claims are made. Is this then an appropriate criterion by which to judge qualitative research? Schwandt (2003) argues not, on the basis that social phenomena are complex and context-specific. Meanings are not context free and cannot therefore be extrapolated beyond the context of a study. However, qualitative research may aim for a degree of 'transferability', i.e. that the findings may be extended beyond the immediate investigative setting through application to other contexts on a case-to-case basis (Malterud, 2001; Sin, 2010; Tobin & Begley, 2004). This is something I considered in s.4.6.4.

4.8.4 Dependability

Reliability relates to the extent to which findings can be replicated. True replication of qualitative research is difficult if not impossible to achieve because of the instability of the social world and the contextualised nature of the research (Sin, 2010). Qualitative studies may exhibit elements of commonality but vary in approaches to data collection and analysis. The process is not standardised. The researcher is the main instrument of data collection and analysis, and the approach taken depends upon the particular researcher's own predilections (Bryman, 2015). It follows that researchers may do things differently, may concentrate on different aspects and may arrive at different conclusions. In addition, researcher and participant co-construct qualitative interviews (Brinkmann & Kvale, 2015). Semi-structured interviews offer flexibility meaning no two interviews are necessarily alike (Marshall & Rossman, 2016). Responses will be affected by the interaction; hence, different researchers will generate different interactions co-constructing different accounts. This too works against replicability and hence, reliability.

For dependability, the process of qualitative research should be "logical, traceable and clearly documented" (Nowell et al., 2017, p.3). It frequently is not (Bryman & Burgess, 1994). In order to establish dependability, paralleling reliability, Lincoln and Guba (1985) suggest the researcher should keep complete and accessible records, including a detailed audit trail at every stage of the study from formulating the research questions through to sampling participants, interviewing, data analysis and drawing-up conclusions. This adds transparency to the process of developing themes, negating the need for replicability in the traditional sense. Following also the recommendation of Guba and Lincoln (1981) for inclusion of verification schemes, the integrity of the research process and the conclusions so developed can also be audited by peers (Bryman, 2015). Mine is a doctoral research study and is necessarily subject to a degree of auditing through internal supervisory mechanisms and progress reviews, and external mechanisms including conference discussions and formal examination.

4.8.5 Confirmability

Confirmability involves showing how the researcher has interpreted the data and reached corroborated conclusions, recognising that participants have put their trust in the researcher's integrity to portray their perspectives (Bell, 2014; Tobin & Begley, 2004).

'Bracketing' requires the researcher to set aside any relevant views and preconceptions before tackling the data premised on the concern that these may influence and shape the analysis (Tufford & Newman, 2012). However, effective bracketing is thought to be virtually impossible and hence, data interpretation will inevitably be subject to a degree of partiality. However, this does not preclude a lone researcher conducting high quality qualitative research, to include the present thesis, which as doctoral research is of a solitary nature, if suitable precautions are taken (Rands & Gansemer-Topf, 2016). In any event, the problem as regards objectivity is not that subjectivity exists but rather when its influence is ignored and/or left unaddressed (Lincoln & Guba, 1985; Malterud, 2001). A commitment to reflexivity throughout the research process is important to addressing this. I included my thoughts on this in Chapter 1, reflexively setting out my philosophical positioning and assumptions, and highlighting any preconceptions and potential biases that could influence this research (see s.1.3.1).

Confirmability may also entail obtaining feedback from participants to assess if I have accurately interpreted and understood their experiences (Sin, 2010). I claim partial compliance in this, having taken great care in accurately transcribing interviews and having sent transcripts to participants for comment. The difficulty and my point of departure comes in not seeking verification of the findings. These derive from my interpretations of interview data whereby individual utterances are collated, decontextualized, re-contextualised and ultimately reduced to form themes. How faithfully my themes describe participants' experiences is mediated through my interaction with the data, limited by my linguistic abilities (Koole, 2012). My themes cannot therefore be subject to verification by individual participants.

4.9 Conclusion

Social constructionism holds that the social world is not a fixed, objective, observable reality; rather it is social actors who create it through their social practices (Cohen, Duberley & Mallon, 2004). Qualitative research draws on the hermeneutic tradition and is thus concerned with subjectivity and with locating socially situated and contextualised meanings (Guba & Lincoln, 2005). Characteristic of qualitative research within a social constructionist paradigm, this study is exploratory, experiential and qualitative, focussed on accessing the socially situated contextualised accounts of human participants with respect to a social practice.

Interviewing is an accepted way of accessing human perceptions, experiences, meanings and realities (Brinkmann & Kvale, 2015; Joffe, 2012). I sampled twenty-one participants from across the five subject disciplines of the host business school, drawing from a pool of academics who could offer a spread of relevant experiences. This followed two pilot interviews that helped with refining the main data collection process. Interviews took semi-structured form comprised of a series of open questions with unstructured prompts. This format is especially common in qualitative social research because it is particularly suited to gathering data on participants' perspectives in their own words (Matthews & Ross, 2010). Two mediating artefacts supported the interviews: firstly, a pre-interview questionnaire completed in advance by the majority of participants; and secondly, two recently assessed undergraduate dissertations professedly exhibiting differing levels of critical thinking. These artefacts facilitated reflection before and during the interview to encourage meta-awareness in participants and elicit insights into the essence of their experiences. I transcribed the interviews, some in full, others by way of comprehensive review, to assist with familiarisation as the first step in performing detailed and systematic thematic analysis leading ultimately to the construction of themes following the six-phase framework of Braun and Clarke (2006).

I have explained the attendant limitations, and ethical considerations, and the steps taken to mitigate or meet these. That this is qualitative research obviates the application of validity and reliability as traditional measures of research rigour. In preference I have detailed the issues concerned with and the steps taken to establish the rigour of this research through the application of 'trustworthiness' criteria applicable to qualitative research (Guba & Lincoln, 1994; Lincoln & Guba, 1985). Ultimately, this study is legitimised through my integrity and honesty (Sin, 2010) and the detailed processes as documented through which I formed and executed this research.

In the two chapters that follow, Chapters 5 and 6, I set out my findings from the thematic analysis. In reporting only two themes, I have strived to develop an optimal and parsimonious number of themes that together capture the key aspects and essence of the dataset. In each interview, I asked participants to describe what critical thinking is for them. I sought to explore this firstly, in general terms; secondly, with reference to undergraduate dissertations; and thirdly, more specifically with reference to sample undergraduate dissertations which for participants would demonstrate disparities in critical thinking. Discussions here fed into Theme 1 (refer Chapter 5). In each interview, I further asked participants to describe how they go about assessing undergraduate dissertations, discussing individual approaches, factors of importance and institutional processes. Discussions here fed into Theme 2 (refer Chapter 6).

5. Theme 1 – Facets of an undergraduate student as critical thinker

5.1 Introduction

Chapter 5 is the first of two chapters in which I report my findings derived from thematic analysis of the data as described in Chapter 4. Here, I develop Theme 1 – Facets of an undergraduate student as critical thinker.

Questions 5-8 of the interview schedule looked to draw out participants' definitions, understandings and conceptualisations of critical thinking. Whilst more often than not responses characterised critical thinking with particular focus on undergraduate dissertations, some also made references that are more direct. Collectively this enabled me to build a picture of business school undergraduate students who are critical thinkers, as perceived by participants, which I have condensed into facets comprised of one key composite skillset, that of argumentation (encompassing structural and evidential aspects); dispositions; and aspects of originality.

5.2 Skills of argumentation

Eighteen of twenty-one participants collectively acknowledged that students should possess a developed skillset that enables them to argue well:

Theodore: *I think to present a critical picture you need to argue a particular point.*

Olivia: *…argument is important because it brings you on board as a reader, and to be able to buy into what the study is about, why it's important, the interpretation of the results, all depends upon the quality of the argument that's being put forward. So I think it's very, very important.*

These quotations from Theodore and Olivia appear representative of the group and come from two highly experienced members of faculty. Both are Associate Professors but on different tracks. Theodore is on teaching track, his role focussed on management and teaching; Olivia on research track, her role focussed on research and teaching. They had fifteen and twenty-one years' experience in academia respectively at the time of interview. Regardless, the message is the same, i.e. the skills of argumentation are fundamental to students who engage with and demonstrate critical thinking.

There are several aspects to exactly what for participants would make for good arguments. I have grouped these under the headings of 'structure' (s.5.2.1) and 'evidence' (s.5.2.2).

5.2.1 Structure

The first area of argumentation that came out strongly in discussions is that students who are thinking critically should be able to construct appropriately structured arguments, i.e. valid and believable arguments that avoid contradictions and which build from legitimate premises through to sensible and substantiated conclusions:

Zhao: *It's about how they actually engage with the arguments. So I do care about not only the contents but also how the arguments are structured.*

Of the participants who discussed argument specifically, the majority considered how this operates at the level of individual arguments within a piece and at the composite level of a dissertation as a whole. The latter should not present as "*a collection of chunks*" (*Hanna*), rather it should present as coherent and logically structured. Structure helps to bring out and make accessible the depth and perceptiveness of students' arguments. However, structuring arguments is something that students often have difficulty with:

Aidan: *…you'll be surprised how many students basically struggle to understand how these components have to be sequenced, and they lose sight of the broader view of what goes into structuring a research argument which is a dissertation.*

If this structure is lacking or deficient, then the messages that students want to convey are at best diluted or at worst lost altogether:

Fiona: *If a dissertation doesn't flow well, if it's not linked together, then it's not really conveying the message convincingly.*

References to logic featured heavily where participants discussed argument during the course of interviews. Logic seemingly provides the spine, enabling an argument to flow and making it work:

Isaac: *I expect to see critical analysis in the mid to high 2:1s, but it's that depth, that perceptiveness, that logical structure to the argument that really makes the difference.*

Logic provides the structure to a dissertation as a whole and its constituent arguments. Reasoning is the glue that holds it all together, signposting the route from the initial formulation of the research through to its conclusions:

Evelyn: *And what we both liked was this was not without its challenges, this was messy, but what she wrote up was not messy and you could follow the logic of how you got there…* **[This came from the document elicitation part of the interview, in which Evelyn referred here to an example dissertation that as described demonstrated a logical structure]**

Evelyn: *…so in there for me it's partly about structure, so there isn't a structure to the logic that's robust enough to defend the conclusions, and in places it was kind of hard to wonder where the hell it came from in the first, you're sort of reading going I just, there are some seismic jumps from here to there with you going but actually you haven't, without the evidence to back it up.* **[This also came from the document elicitation part of the interview, in which Evelyn referred here to an example dissertation that as described contrastingly lacked a logical structure]**

This comparison by Evelyn of two dissertations shows the worth of structure. With the former, Evelyn as supervisor was clearly aware of the difficulties faced by the student in pulling together a challenging and quite "*messy*" (*Evelyn*) piece of research, but was impressed at just how well the student was able to convey the overall argument because it flowed logically. The latter, in contrast, appeared to lack sufficient structure so that it did not flow logically, with the conclusions not necessarily following from or properly drawn from the premises. Evidently, that the overall argument progresses logically is important to establishing the credibility of any conclusions reached and claims made:

Zhao: *Critical thinking should be in the form of arguments at the surface but we want to see how good the arguments are. I mentioned logic, that means logic it's a way of thinking but in the piece of work it's how the student actually expresses his or her view. The logic has to be smooth flow. It represents how well the student actually thinks, argues.*

If the logic and reasoning that justify the conclusions are clear and evident then this will go some way to satisfying the reader who can then track the progress of the argument:

Ella: *It [logical argument] has to be, I think it has to be, yes I think if it's just really unstructured so you do not have that clear pathway of how did you come up to, so I need to be able to find out how ideas formed and then how they made up their statement or conclusion.*

Viola, an economist with a background in quantitative research that involves the construction of mathematical models of economic phenomena, equates logic with a scientific approach to research:

Viola: *… you have to link and be able to interpret that logic that I was saying before. It's fundamental, otherwise it's not science. We follow the scientific methods, the mathematical methods. Otherwise, it's not science.*

In economic modelling you have assumptions, and within these assumptions you build a model through which you generate results, which you then criticise in terms of the assumptions. Mathematics underpins economics as a discipline and provides a ready-made 'scientific' logic for Viola's research:

Viola: *And this is how mathematics works as well, right, when you prove a theorem and so on? So yes that helps a lot. You look like that, given these assumptions you find these results, but if you relax the assumptions something else might come up. They understand that it is this process. I will say, yes, in this mathematical logic helps a lot and it clarifies lots of things. So that's an important tool that we may use.*

Evelyn, an accountant with an undergraduate degree in chemistry and a background in qualitative research, also equates logic with a scientific approach:

Evelyn: *So I am a chemist first and then I'm a chartered accountant, both of which are disciplines of rigour in thought and logical progression and making sure that what you claim at the end of it is absolutely what you've got firm evidence for and you don't over claim, if that makes sense.*

For Evelyn, analytical thinking in science is much the same as critical thinking in accounting and other business-related disciplinary contexts:

Evelyn: *…it was only when I came in to academia did I realise that I did it [critical thinking] anyway. So I'd never called it critical thinking. I thought it was being logical, that's what I thought it was.*

Viola and Evelyn are two highly experienced members of faculty, with nineteen and twenty-four years in academia respectively. They come from different disciplines, economics and accounting respectively, and have different academic backgrounds, yet both are of the fundamental view that a scientific approach, and the logic inherent to this, are important for argumentation and hence demonstrating critical thinking.

In summary, participants view structure as important to effective argumentation and the rules of logic provide the necessary structure.

5.2.2 Evidence

The second area of argumentation that came out strongly in discussions is that students need to win the argument on paper as it were. To do so they need to substantiate their (logically structured) arguments with evidence. The credibility of the arguments put forth will diminish if students are unable to offer suitable supporting evidence for their conclusions and thereby convince the reader of the efficacy of their arguments:

Arthur: *They have to put together an argument and draw on evidence, find different bits of literature that support their view…*

Florence: *…throughout I expect students to be able to present a well-referenced, supported argument. I think the whole credibility of an argument can be let down if the student isn't actually able to provide suitable and relevant referencing to help support the points they are trying to make.*

Arguments need evidence. Conclusions need backing. In part, as Florence suggests, evidence takes the form of referencing appropriate academic sources. Simply referencing sources is not sufficient, however. Half of the participants who discussed this advised that evidencing skilled argumentation requires the ability to engage with and evidence 'critique' of sources:

Evelyn: *…really critique and evidence that critical thought…*

It would seem that students are expected to add weight to their arguments, displaying their critical thinking through not only referencing sources but also critique of the views and arguments therein. Participants' comments here were generally contextualised to undergraduate dissertations with, according to approximately half of commenting participants, students expected to evidence that they have engaged in detailed evaluative reviews of the extant literature on their research topics through critique of published academic works. Here, students must demonstrate their understanding of the state of knowledge on their research topic. Hence, they have to read. They have to gain an initial understanding of what has been done and they have to acknowledge the existing work as a starter. They should not stop there, however. Critique, incorporating interpretation, analysis, evaluation and inference, must progress beyond simply describing the views of others. Most commenting participants discussed the importance of students comparing and contrasting sources, for example, suggesting that this would evidence critical thinking and so be worth crediting:

Oliver: *I would still be willing to give a first class mark if it was simply comparing and contrasting the views of authors.*

The majority of participants, Oliver included, expect to find that students have compared and contrasted sources in the literature review part of the dissertation. The corollary to this is that the absence of comparing and contrasting of sources would present as 'descriptive' rather than critical (refer to s.5.4.3 for a discussion of what 'descriptive' means for participants and why this is viewed as uncritical). For Oliver, an accounting academic of forty years' experience, comparing and contrasting sources in itself could evidence a level of critical thinking commensurate with a first class grade; at least as far as the literature review aspect of the dissertation is concerned. Whereas for the bulk of participants this by itself is unlikely to be enough. There seems to be a graduation in critical thinking from acknowledging sources through to comparing and contrasting sources through to critique and synthesis of the literature. Comparing and contrasting is a necessary step, a foundation, a rung on the critical thinking ladder, but not one that takes you to the top:

Oliver: *…if a dissertation was more of a narrative description of the views of the established authors without at least an attempt at a critique of those authors then I wouldn't give it a first class mark.*

Here now Oliver is explaining that critique of sources is clearly important to attaining a first class grade. At first glance this seems at odds with his prior comment that he "*would still be willing to give a first class mark if it was simply comparing and contrasting the views of authors*" (*Oliver*). Working on the premise that critique is more than comparing and contrasting, it is possible that Oliver was contradictory here. An alternative interpretation is that for Oliver 'comparing and contrasting' and 'critique' are synonymous. However, I do perceive a difference in these terms across participants, with the latter building on the former. Students who are able to critique are not simply acknowledging sources, not just describing them, not uncritically accepting of sources as representative of the current state of knowledge. They are making links between them, comparing and contrasting points, unpicking arguments, examining areas relative strengths and weaknesses, evaluating findings, identifying tensions and questioning apparent or accepted positions, i.e. students who are thinking critically will critique the arguments put forth by sources and judge whether their findings should be believed:

Aidan: *Because you are critiquing, right? You're critiquing the state of the knowledge and this is because you have thought through the logic of it and intuitively, or from your past experience, you know that this is insufficient, or it can actually do better. So that that's what it means to be critical. Critical means non-accepting of what is the current state of the knowledge.*

The literature review provides students with an opportunity to synthesise all of this critique. Students who take this opportunity are then showing that they have not just evaluated individual papers in isolation but importantly have synthesised knowledge across the literature base, making links and finding synergies:

Olivia: *…I would want to see at least some synthesis of the literature as a basic critical skill. So not just listing different bits of literature but being able to synthesise it and put it together in a way.*

Fiona: *…it's the difference between being able to synthesise academic literature rather than having a paragraph about each article connected together. That's one skill which is to my mind very good and demonstrates critical thinking.*

Olivia offered an interesting metaphor to explain this idea of synthesis:

Olivia: *I always use the metaphor of a jigsaw puzzle and a literature review is a jigsaw puzzle. And you've got the box, but you don't have a picture on the top, so you've got all these pieces and you need to figure out how they fit together and what picture that it makes. And it may be that the picture has bits missing or one piece that fits there but then there's a whole bunch of pieces that link together over here. What their literature review is, is to try and put that picture into words. To describe if this is the topic what does it look like?*

It seems that students who are thinking critically are able to synthesise the picture from its parts and put this into words. To do that they need to understand what the different pieces are, how they relate to one another and how they link together. The pieces are not just discrete. In some cases, studies can appear disparate but there are relationships between them to be uncovered. If students can figure out what those relationships are then they can start to make sense of the literature and synthesise this into a coherent picture.

This issue of synthesis applies similarly to students' developing their own findings from their research:

Olivia: *…you shouldn't just describe the data because just describing the data, to be honest, anyone could do that with a vague, no subject or not. You're trying to provide a different reading of the data because you have developed a knowledge and an understanding of a concept, so you'll look at that in a different way.*

Credible positions are defensible. Poorly explained findings, misunderstandings or errors, unsubstantiated assumptions and claims will all inevitably detract:

Hanna: *I say well you need to be careful that in the end you don't end up with a conclusion that you cannot back up. You cannot say all the evidence points in this direction or the theories point in this direction and then I will say I don't care, I'm going to be Donald Trump. That shouldn't happen.*

Absent synthesis of findings and literature, what may come across is conjecture, students' own opinions, potentially ill thought-through, lacking in evidential support and hence lacking in critical thinking. Students should demonstrate that they have used the literature base as a lens through which to interpret the data and as a means of providing evidential support:

Zhao: *This is definitely what we expect at this level. They are not just accepting knowledge. They are thinking. They are using the current bundle of knowledge perhaps to figure out their own opinion based on the literature.*

Oscar: *So there's no links back to the literature but more worryingly, or I would say or why this is less good, is that it's quite descriptive. It doesn't try and pick up a concept or a theory that might help explain this. It just says this is the way it is, these are all of these motives. Well ok but is there not a theory that might try and explain this and then think a bit more broadly beyond that?* **[This came from the document elicitation part of the interview, in which Oscar referred here to an example dissertation that he chose as being evidentially low in critical thinking]**

Self-critique is important too in this context. Critique need not be limited to an examination of the works and views of others but should encompass critique by students of their own findings:

Milo: *And then when they're coming up with results, there's just this idea that it's an ongoing discussion and this is what I have found and this is how my finding is different or is the same or is, and again it's the idea that they're critiquing not only those studies that have gone before but also their own study.*

At one level, students can show where their analysis and interpretation of the data is consistent with evidence from the literature. Better dissertations will show where these are consistent and inconsistent. Better dissertations still will tease out how and why consistencies and inconsistencies exist:

Evelyn: *…why in my data would I get a different perspective from what is going on here or why is my data different from the pre-existing literature and yet stills gives us the same answer?*

Therefore, there should not just be a simple presentation of results but a deeper engagement considering why the findings are what they are, and how they compare to expectations drawn from the literature. This being the case it really should not matter, or it should matter much less, whether the assessor actually agrees or not with the student's point of view. What should matter is that the student has supported his or her points of interpretation, in particular with reference to and application of the synthesised literature base:

Olivia: *It doesn't necessarily have to be my view, because very often I'm quite happy for them to have a different perspective, but as long as that's why I like you and that's convincing, and that use of the data to do that and they've supported their interpretations through the literature, then I think that's what we're looking for.*

In summary, students who are thinking critically will produce arguments that are appropriately structured and supported by suitable evidence. This relies not only upon acknowledgement, critique and synthesis of sources, but also upon adroit synthesis of findings with the literature base.

5.3 Dispositions

5.3.1 Scepticism

"Scepticism is great doubt about whether something is true or useful." (Collins English Dictionary Online, 2020).

That critical thinking is underpinned by a propensity for scepticism, a willingness to doubt towards authorities and received wisdom, was highlighted by thirteen of twenty-one participants. Scepticism then is viewed as significant to the ability of students to think critically:

Harry: *The natural sciences don't really, they have a hefty scepticism, so we should have the same at least in social sciences.*

Universities ought to be facilitating the development of 'sceptical' students who as they graduate can go out into the world and ask questions of things. Regrettably, however, at least for Viola, such scepticism is not readily apparent from undergraduate students' dissertations:

Viola: *I find that they accept, most of my students accept.*

Viola seems to feel that many of her students are all too accepting of what they read and that they fail to question sufficiently, apparently lacking the necessary scepticism. Presumably then evidence of scepticism, if desired but purportedly rare, would serve to mark out the higher performing, more critical students. Students who are sceptical demonstrate critical thinking because they do not just take things for granted or as given but are inquisitive, they ask questions, they doubt and they probe to uncover how and why things are the way they are:

Mason: *… what's my definition if you like of critical thinking? Not taking things at face value…*

Olivia: *I think that those are stuff that if you've got an inquiring or questioning mind, that you will look at the world in a different, and you'll start to ask questions.*

Participants invariably saw scepticism in a positive light:

Aidan: *…you might develop a cynical attitude, but the fact that you develop a cynical attitude means you actually have been thinking about some things otherwise you're not able to then challenge how the world is, right, in order to develop that cynicism.*

Aiden talks of "*cynicism*" here. Cynicism is listed as a synonym for scepticism (Collins English Thesaurus Online, 2020). However, cynicism and scepticism may be distinguished because the former has negative overtones whereas the latter does not. Scepticism here is viewed very much in a positive sense. Thus, students who are critical thinkers should possess a healthy scepticism, the habit of mind to doubt, to question, to challenge, to strive for meanings and justifications. With this mind-set comes a reticence to accept the current state of knowledge and importantly a willingness and a confidence to challenge received wisdom:

Isaac: *Yes, it would be being inquisitive. The text book says this, it's been in the text books for years. Well, I'm not just going to accept that. I want to have a look at the evidence. I want to challenge that if it needs challenging.*

Reuben: *Now critical thinking is when they stop and you can see the cogs going around. So that's why critical thinking is to me looking at alternative explanations and not just taking things on that face value.*

Sceptical students are also not simply accepting of orthodoxy. They think it all through, using the current bundle of what is known to figure things out for themselves. The scope of scepticism is not, however, limited to challenging knowledge, cynically or otherwise. This is an important aspect in context but the sceptical student moves beyond this, opening up and seeing the world not just in a more questioning way but also for Reuben in a more personal way:

Reuben: *I've given you an example of that in the student just was sitting on my module and said well I just totally disagree with this because that's not my real lived experience. And then what she did was she took her real lived experience of working in, shall we say a low cost supermarket chain, and said well they don't have inspirational leaders they have very directive leaders, and she explored that through the course material she was getting. I thought that was a great example of critical thinking.*

Scepticism may also be revealed in students being alive to and aware of the limitations of their research. Sceptical students will acknowledge these and be constructively self-critical, exercising suitable caution with respect to any claims made in light of the limitations whilst recognising that it is possible for alternative explanations to exist:

Mason: *So again, going back to the criticality, it's all part of being critical. It's not only not accepting anything at face value, which in a sense includes your own work, think of being critical of that.*

Penelope: *And then they've thought quite deeply about what the analysis is telling them in relation to the question, but they're not afraid to say actually there's limitations here, there's gaps here and that they've suspended judgement…*

5.3.2 Open-mindedness

"If you describe someone as open-minded, you approve of them because they are willing to listen to and consider other people's ideas and suggestions." (Collins English Dictionary Online, 2020)

Fourteen of twenty-one participants referred to what in combination I have interpreted as the need for critically thinking students to be 'open-minded'. Students who are open-minded are curious and inquisitive, are appreciative of the fact that they do not know everything, are able to recognise what they do not know and are accepting of the need to work on this, whilst open objectively to different possibilities:

Reuben: *It's almost that critical thinking is a journey really, you don't know what you don't know and sometimes it's, critical thinking to me is trying to find out what you don't know, that's the real thing about really trying to turn the rocks over and see what's there...*

Penelope: *An openness and not taking everything at face value. Yes, being open to look at things at a deeper level, an attitude of being open to learning more, going into things in more depth…*

Participants plainly look for students to adopt a position and argue effectively for this. Evidently, any position taken by students needs justifying. It requires alternative views to be considered, showing an awareness that people may have different views. Students are not expected to adjudicate in the debates necessarily but should make an informed judgement as to where they feel the power and efficacy lie, recognising where views oppose and offering evidence to support and/or contradict those views, and showing that they understand the theoretical, methodological and contextual perspectives that underpin these in coming to their own positions:

Theodore: *So a well-reasoned argument is something where I think the student would carefully articulate why they are taking a particular position, and they would go through the pros and cons and present a good case for what they are saying.*

Positionality then is about students being able to work through multiple perspectives, addressing the positions taken by others, recognising that there is not necessarily a right way of interpreting something, and ultimately developing and making clear cases for their own positions. Being open to different interpretations, different perspectives, different approaches, enables open-minded students to have greater appreciation of context and a deeper understanding of the positions they ultimately adopt. This requires openness yes, but also fair-mindedness and disciplined thinking. Open-minded students are not impulsive or precipitate in taking positions; they think it all through and arrive at reasoned determinations:

Olivia: *But that blind kind of no this is my position, I don't care what anyone else says, that doesn't demonstrate a level of inquiry and open-mindedness and curiosity. It's just blind if that makes sense? Oh this is my way of seeing it and there can't be another way.*

Olivia: *What I would expect to see in a critically informed approach to doing that would be a consideration of different perspectives, but a strong case made for the perspective that's being pursued, rather than a blind let's ignore everything that doesn't fit with the position that I want to take and just make this particular case.*

One further aspect of open-mindedness, which came up in conversation with around a third of participants, was the idea that students should be able to see beyond what is in front of them. Beyond, for example, the immediate context and narrow confines of their own research through to greater concerns. Open-minded students can "*see the big picture*" (*Fiona*). For example, they can place their studies into the larger portrait of which they form just a part:

Fiona: *…there are people who can see the big picture and relate different things to each other, and there are people who can't do that very easily, they can't see that big picture, and I guess that's a different dimension of critical thinking that I've become more aware of…*

This is not to suggest that open-minded students must be ‘blue sky thinkers’ necessarily, rather that students should be sufficiently open-minded to show that they can think through the wider issues and implications of their research and can determine where that fits into the greater overall research landscape:

Evelyn: *…there must be an ability to see the wider implications of what’s going on and to make those sideways steps that go actually I was looking for this but just look at that, hang on a minute that matters.*

5.3.3 Flexibility

“Something or someone that is flexible is able to change easily and adapt to different conditions and circumstances as they occur.” (Collins English Dictionary Online, 2020).

With critical thinking comes the flexibility needed to ‘get the job done right’. Students can also demonstrate flexibility in that having sifted multiple perspectives, and weighed up the evidence, they have the presence of mind to self-correct, to change their views and reform any position they have taken in light of fresh perspectives and new evidence:

Evelyn: *Logic and robustness and open mindedness I always thought that, but then there was a phrase in my Cert Ed which said critical analysis or critical thinking is knowing why you think what you think, what assumptions there are behind that and therefore if you are shown that your assumptions are wrong you are able to change your mind.*

Conducting research is seldom without its trials and tribulations. Students who think critically can demonstrate their flexibility through the dissertation process. For example, flexible students are persistent; they are unfazed by problems in having to reform their research questions or methodologies midway through the process where issues present and adjustments are necessary. Academics recognise that the research process is “*messy*” (*Evelyn*) and seldom runs straight and true. Students who are flexible are upfront about this and demonstrate how they have dealt with the pitfalls along the way:

Arthur: *Flag up your own weaknesses. You won’t lose marks. Suggest a solution, you might gain a few in the process. If you point out your own weaknesses and suggest a solution you can sometimes end up, depending how well you do it, with more marks than if it had just worked perfectly.*

Students can thus demonstrate flexibility and hence critical thinking by recognising the limitations of their research, showing that they have given due consideration to what went wrong and why but also what action they took in response. That is a rich part of the research process and students should feel comfortable in reporting this, being critical of their own research approaches and adaptable in finding solutions to problems. Some things simply do not work as anticipated or planned. That is the nature of research. Academics know this. Honesty and flexibility are seen as positives:

Oscar: *I quite like where there is an element of reflexivity about a struggle that has gone on, particularly with some projects. It shows a good appreciation of the research process. Some things don't work but it doesn't mean they're wrong it just means that they haven't worked, but to then pretend that they didn't exist and everything is you know very smooth is not being truthful to the research process.*

5.4 Originality

Originality is defined as "the quality or condition of being original", or "the ability to create or innovate" (Collins English Dictionary Online, 2020), or "the quality of being special and interesting and not the same as anything else" (Cambridge Dictionary online, 2020).

Seventeen of twenty-one participants raised issues and aspects of originality in connection with critical thinking in undergraduate dissertations. In particular, participants discussed this in terms of what for them originality is, i.e. creativity, making a contribution and original thinking, but not 'descriptive' accounts or 'replication-style' studies.

5.4.1 Creativity

That thinking critically involves thinking creatively was discussed by twelve of twenty-one participants. Creativity, in this context, relates to students thinking differently about their research topics and the problems at hand, evidenced perhaps by going beyond the norm in their approaches to research or in their thinking:

Zhao: *Like what we ask our students to do we have to push the boundary and perhaps into another space, so that's creativity…*

Viola: *Their thought might be more beyond the curriculum, which is always very, very nice, seeing a student who is engaged and there is creativity.*

Zhao is I think suggesting that students should go beyond the mundane and formulaic but how does this lend itself practically to research? How exactly can students evidence that they have pushed the boundaries? How can they demonstrate that they have in effect shown some creativity? It would seem that one of the main ways of showing creativity is for students to be able to recognise what is new, what is novel, and that they are able to bring some element of this into their research:

Fiona: *It can be a really novel idea. It can be a bold way of doing something.*

Zhao: *…we simply talk about novelty. Being novel means that you have to be critical.*

Viola: *I think it is part of being critical, exactly, because without being critical you cannot recognise what is niche, what is new.*

Novel/novelty/niche – this can come anywhere in the dissertation, including its context, subject and methods. The setting of research objectives and/or questions (hereafter referred to as questions for brevity) to frame a dissertation study is one area in which students can demonstrate their creativity. This is not something that students can necessarily get from a textbook. Yes, a book can detail what a critical question might look like, but creative students will be asking themselves different types of questions, inquisitively pushing for something, developing imaginative questions as a way into and of progressing their research. Creativity then can in practical terms with respect to dissertations manifest in the novelty of research questions:

Viola: *So the research question is very important. It must be somehow engaging for them and niche. I like that. Their thought might be more beyond maybe the curriculum, which is always very, very nice seeing a student who is engaged and there is creativity. I like that.*

However, although participants may desire creativity in research questions it is perhaps less often seen, to the detriment of the piece:

Olga: *I know that they will be using certain data analysis methods but the research questions that they choose tend to be from a very limited range…maybe it is unfair to expect from them, given what they learn in the programme, to come up with a very different topic. But then this for me is a lack of critical thinking because they just take what they…low-hanging fruit.*

That said, whilst creativity can manifest in the creativity of the questions, the lack thereof does not necessarily mean that the piece will lack critical thinking. This could equally be apparent in the creativity with which the questions are pursued. Even a study based on a well-trodden path with familiar research questions can be interesting and exhibit creativity in the way in which the student has gone about fulfilling these:

Aidan: *…but they have then a set of theories up front, so in that case I think critical there means you know how you're going to interpret your results given that they're different and what are the kind of questions that you will need to ask. So it's being able to show that you're able to ask intelligent questions and I think there's a creative element to that, you know?*

5.4.2 Contribution

The majority of the participants who commented upon originality were of the opinion that an undergraduate dissertation should contribute in some way:

Olivia: *I am looking for something that makes me see something in a new way. So there has to be some kind of contribution…*

Aidan: *I think if you if you if you're not trying to push the boundaries even if it's a small boundary [inaudible] to look beyond what has already been done then you know it's very hard to say then you have critical thinking…*

It seems that at least some benefit has to be derived from the research. There has to be a point to it and an outcome that adds in some way. I took this to mean a contribution to knowledge. This is a known requirement, at least at the level of peer-reviewed journal articles for which it is a *de facto* condition of publication; an aspect that reviewers obviously look for. Similarly for doctoral theses and their examiners.

Nine participants discussed having developed their own views of critical thinking from engaging in their own research. From this, they learned to interrogate literature, ideas, theories and data in a way that helps develop understandings that are more critical. They further benefit from being part of research communities made up of academics who have different ways of thinking about things, who can freely articulate their thinking and who can engage in conversations through which learned discussion can and does take place leading to the further development of knowledge:

Aiden: *I think you have to do research. This is why research, if you want to use the kind of trite university slogan of research-led education, I think the more research you do the more you understand what is the process of knowledge.*

This involves contact between academics within and without institutions, at conferences for example, but also through submission of articles to peer-reviewed journals for publication. Journal reviewers demand the highest levels of critical thinking for publication:

Zhao: *I would say it's getting harder and harder to get published. It's actually for the top journals which requires everything: theory, knowledge advancement, novelty, unique datasets and appropriate methodology and interesting and sometimes and counter-intuitive findings, that's everything. I learned out of it it's the critical thinking.*

There is then evident pressure to demonstrate critical thinking in submissions. This is something that editors and reviewers expect and will readily comment upon:

Oliver: *… when we write something ourselves referees will read it and perhaps savage it, so we are still being assessed on our own critical thinking.*

Participants confirm that they learn from this process of review and feedback. Coupled with experience, with time and with practice, this helps them to develop their views of critical thinking and thereby improve the standard of their submissions:

Reuben: *And I think for me the more research I've done, the more I publish, the more I get rejected from journals, which doesn't do me any good but then you read them after the emotion is gone and it's really helpful to see the peer review of your work.*

Mason: *I mean the standards that you apply, or the standards that I have applied throughout my career, you know I look back at some of those early papers and think oh my goodness I've no idea why it got…it wasn't what I would now regard as a good piece of work because my standards have changed, my understanding of what is expected has changed.*

Returning now to students' contributions to knowledge, these can be similarly drawn in grand terms, for example, in doing something that has never been done before, or never before in a particular way, or in making some sort of breakthrough that leads to new knowledge. Appreciably, undergraduate students are unlikely to operate at the same level as academics or doctoral students and are unlikely to generate new knowledge from their research. For participants, such a contribution would stand out and surprise because they would not expect it of undergraduate students:

Isaac: *The vast majority of students aren't going to make a breakthrough at that particular point in their academic career. That level of originality, a brand-new idea, you're not going to see those very often at undergraduate levels.*

Florence: *It's the sort of thing that perhaps, I think everyone would expect it to be an excellent piece of work, but if they're then demonstrating an original contribution to knowledge that's what would push it into the outstanding range.*

Much of the discussion in interviews of 'contribution' was geared towards examining what this might constitute in practical terms. It seems that any such contributions from undergraduate students could take different forms and be construed at different levels. For near half of commenting participants, this starts with the literature and an expectation that students will synthesise this to the point of identifying a gap that research can then usefully target. This is one way in which undergraduate students through their dissertations can make a contribution:

Mason: *…I push my students to say what is it that you're doing that we don't know enough about? The literature just doesn't quite cover that area and therefore what contribution are you making? What gap is it that you are seeking to fill?*

Hanna: *I think for the really high, for the 80s, 90s, and if you look at the description for the 90s they basically have to come up with a gap. Well I think you should get an 88 without having a wonderful gap but I think surely the 90s that's just what the marking descriptors say, you've got to come up with something that people don't usually think about.*

Other simple ways of contributing incrementally are suggested. One is to adopt a different tack, for example, looking at a specific known area and bringing in some interesting aspect, or interesting context, or examining an interesting group of people, or applying theory that is not normally applied to that context and so on:

Milo: *Well in that it's not original, it's not that it's something that's coming completely out of their own head original, but is that it's original in that it's adding something to what's in existence already. So in that context it is original but it's not like an original thought. It's like an addition, if you like, or a contribution that adds rather than revolutionises.*

Students are effectively then part of the academic conversation, building upon the work of others and making their own contributions:

Mason: *…yes I'm still looking at it advancing our knowledge in a particularly area, even though it may be reasonably narrow and so forth.*

Milo: *I think some students, the students who are probably best at critical thinking, are able to appreciate that their piece is just part of that [knowledge base]…*

Students may also demonstrate a contribution and hence, originality by recognising and communicating the novelty of their findings, of how and why their findings are special:

Florence: *Students who can go beyond my examples and actually come up with some insights as to why their findings are different, that to me is really an elegant display of critical thinking.*

Insight can come from students demonstrating that they have thought about and through any wider implications of their studies and are then able then to answer the all-important 'so what' question in relation to their findings:

Oscar: *To me the conclusion should think about the wider implications of the research, why anyone should care about it.*

Milo: *But I'd like to see the better students, they think a bit more about maybe what the implications are of their findings.*

Additionally, where students might realistically contribute is in offering suggestions for professional applications or policy implications, because business-related research is often situated in real-world situations in which protagonists are looking for practical solutions to problems not theoretical contributions:

Oliver: *…I am looking for some sort of bridge between the model and then the reality say of the accounting and business world.*

Reuben: *It's got to have some, in business and management, it's got to have some implications for practice, really, is what I suggest for them to think about.*

Milo: *…these are the policy implications of my findings, or implications for industry or investors, or thoughts for further research. I think all of that is critical.*

5.4.3 Original thinking

Two-thirds of participants who discussed originality considered the idea of students demonstrably self-investing in their dissertations, and that this could serve to evidence originality:

Hanna: *You haven't actually told me anything about your own understanding. And I think that then ends up really descriptive and not analytical and not critical.*

In essence, if students own original thinking were to come through their work, evidencing application of their intellect to the problems under investigation, demonstrating their understandings, impressions, perspectives and positions, then that could be perceived as offering some original perspectives and hence present as more critical:

Arthur: *The very best students are showing evidence from that literature review chapter that they're starting to form their own impression.*

The key concern here is whether the student's own points of view should be detectable in the piece. In the main, it seems that participants are interested in what students think and have to say and that this should come through in some way:

Theodore: *The student's own voice I don't think needs to come through in the first person. A dissertation doesn't need to be written in the first person. But I do think that the student's own voice and view and demonstration of what they think about this does need to come through, yes.*

Arthur: *You get the impression of a voice, however raw it might be coming through. That shows that they're thinking about stuff and adopting opinions in that voice, or adopting views and perspectives.* **[This came from the document elicitation part of the interview, in which Arthur referred here to an example dissertation that he chose as being evidentially high in critical thinking]**

This helps signify that students are not just working to a dissertation formula, are not just regurgitating the ideas of others, but are taking ownership of the material, reconstituting it in their own thoughts, forming and articulating their own original ideas:

Evelyn: *…[a critical thinker is] a person who is thinking about it and is sentient, rather than going this is a recipe for doing a dissertation and on I go…*

Oliver: *It is something that stands out, if somebody is not just replicating a model, is not just moving through techniques but is trying to sit back and apply their own thought to what they have read and to try to derive some conclusions…*

5.4.4 Descriptive accounts

The majority of participants who discussed originality referred to 'description' in quite negative terms. It seems this is a well-worn pejorative term denoting the antithesis of originality:

Penelope: *What it [originality] isn't, I know what it isn't, it's not being descriptive, so it's not describing something.*

A descriptive account is perceived to be basic, unsophisticated in research terms, uncreative, would tend to report or describe findings far more than analyse and interpret, would fail to think through the implications and develop the contribution of the study. Oscar offers an example:

Oscar: *Pretty much just a descriptive literature review, some findings that are described and a conclusion that says this is what the findings were. But it then goes no further forward. It's not saying well why have I found this and what informs this? What are the implications of this? It's just description. It's nicely written, there's nice engagement with the literature, but it's very much descriptive, descriptive and then the end.*

[This came from the document elicitation part of the interview, in which Oscar referred here to an example dissertation that he chose as being evidentially low in critical thinking]

Description tends towards the superficial, lacking in depth of thinking, with the student making little attempt to make sense of it all. Description is thus a label best avoided by students who wish to do demonstrate originality and hence, critical thinking:

Mason: *I can't remember now exactly what I said but it would have been along the lines of this is pretty much descriptive. It doesn't really do anything that's critical. You could almost tell just by looking at the contents here. Literature review, page 6. Company introductions, page 8. Ah there's a problem!* **[This came from the document elicitation part of the interview, in which Mason referred here to an example dissertation that he chose as being evidentially low in critical thinking]**

So what does it mean to be descriptive in this context? How does this present for participants? One clear indicator of this is where the student adopts a “furniture sale catalogue” approach to a literature review (Haywood & Wragg, 1982, p.2; cited in Brennan, 1998, p.14). Here, students provide a list of sources in which they simply state that x says this and y says the other, or that this paper does this and finds this and the next paper does that and finds that and so on. The sources may have relevance but students are not explaining how they compare, or how each adds to our knowledge in some way, they are just reporting sources like a shopping list: two carrots, one pack of butter, two pints of milk, etc. It is listing, it is repeating, it is not critique, not demonstrating their knowledge and understanding, and it is forming much less of a holistic view of the literature than would point to critical thinking. Such an approach would also then tend to lack original insight. As a result, this could be construed as uncritical:

Florence: *To me, it would be very difficult to get to a first-class level if say the literature bit was simply X in 2000 said this, Y in 2001 said this, Z in 2002 said that. I would be looking to go up in terms of in-depth analysis, critical evaluation.*

Reuben: *…and this is where we come into the criticality, a really weak dissertation to me would almost be a timeline, a descriptive (INV: chronology?), yes, you know that in 1982 somebody did this and you know and there's no synthesis of the thinking.*

An example from psychology might be a literature review that says Freud did this and then Pavlov came along and said that. If this is reported chronologically without synthesising the sources the student may fail to pick up on the fact that Freud and Pavlov were engaged in debate and did not agree. So that rounded understanding of the literature that is so needed to evidence critical thinking would be lacking.

Moving into their findings and analysis, students have to know about what they are discussing. They have to demonstrate their knowledge and understanding there too but now with regard to their own research. This is in part evidenced through appropriately describing what they have found. However, demonstrating originality, and hence critical thinking, obviously goes beyond first stage description:

Olivia: *…in things like data analysis when you're going beyond just describing he said she said, or they did he did, in like an observation situation, and actually trying to then provide an interpretation of that. Ok so they did that but what is that? What are they doing here? What's going on with that?*

At a descriptive level the student could write simply that the results were this or the results were that, i.e. just describe them as they are. However, participants would expect that students have taken a more evaluative approach to the analysis and interpretation of the data. There is then more than a simple presentation of results but a deeper engagement evident through, for example, an explanation of why the results are what they are, how that compares with any prior expectations and how that compares with knowledge synthesised from the literature base:

Olivia: *So their evaluation would be starting to give in to that higher level, critical thinking. Like when you're really starting to think about well ok so I saw it, I understood it, but what does that really mean in terms of coming back to understanding?*

This then is where the student is engaging in critical thinking, going beyond the simplistic description of results and really thinking about them leading to insights, which brings us back to originality:

Florence: *But the sort of things with critical thinking, I've talked about literature review and what I except from that, but discussion chapters, where they actually think about and offer some really insightful evaluation of why their findings have diverged.*

5.4.5 Replication-style studies

Half of participants who discussed originality commented in this context upon replication-style studies. These are popular in some business-related disciplines, particularly Finance. I have labelled this approach 'replication-style' because it does not involve a straight replication, i.e. repeating a prior study exactly as is. Here, we have a situation where a student takes a published study and attempts to re-perform it, but with some revised aspect such as the use of fresh data, or situating the study in a different context, etc. For example, a student might take a published paper that has investigated herding behaviour in financial markets using data collected from US markets and redo this using data from another context such as European markets, Chinese markets, Indian markets, etc. Alternatively, if the published paper collected data from say 2006-2013, then an updated study could focus on 2014-2020. Both would allow for comparisons to be made and potential insights to be formed, and carry the added advantage for the student of providing a methodological blueprint to follow. All very positive it seems. Nonetheless, there was evident disagreement amongst participants on the worth of this approach, with a small majority in favour.

Aiden provides here an example of disagreement with the approach:

Aidan: *I think if you if you if you're not trying to push the boundaries even if it's a small boundary [inaudible] to look beyond what has already been done then you know it's very hard to say then you have critical thinking, because then all you're doing is you're doing a kind of like a replication study or somebody has given you the idea and you're just basically applying it.*

This seems to suggest that a replication-style approach inherently lacks originality, that what is sought from students is more than simply the (unoriginal) repetition of techniques or the 'cranking of a lever' to operate a model to produce a different set of numbers based on a different dataset:

Fiona: *Oh a boring method, content analysis using the GRI framework that everyone's used. So, this is something that has been done so many times before and written so many times before in a similar way.* **[This came from the document elicitation part of the interview, in which Fiona referred here to an example dissertation that she chose as being evidentially low in critical thinking]**

Florence: *I think it does demonstrate a higher level of critical thinking if they do actually go for something that's not just a replication study. I think if they can expand that, at the very least add to it, integrate two models together. People have tested these hypotheses before but not as part of a broader framework or as a broader model. And I think that's what demonstrates the critical thinking.* **[This came from the document elicitation part of the interview, in which Florence referred here to an example dissertation that she chose as being evidentially high in critical thinking]**

More positively, a replication-style approach is perceived to be less risky for students. A safer option then, albeit compromised, for busy time-limited undergraduates. However, the support here for this type of approach was for the avoidance of risk. In no way did participants categorise a replication-style approach as offering originality.

5.5 Theme 1: Summary

Theme 1 presents a conceptualisation of critical thinking in the host business school, interpreted from the accounts of participating academics. This is comprised of skills (of argumentation); three headline dispositions (scepticism, open-mindedness and flexibility); and originality; simply represented in the following diagram as three corners of an isosceles triangle, with skills and dispositions forming the base and originality at the top. The dotted line signifies that whilst participants generally agreed upon skills and dispositions at the base of the triangle, the inclusion of originality at the level of the undergraduate student is disputed (see s.6.2.2 for a fuller account of this issue) (Fig.5.1):

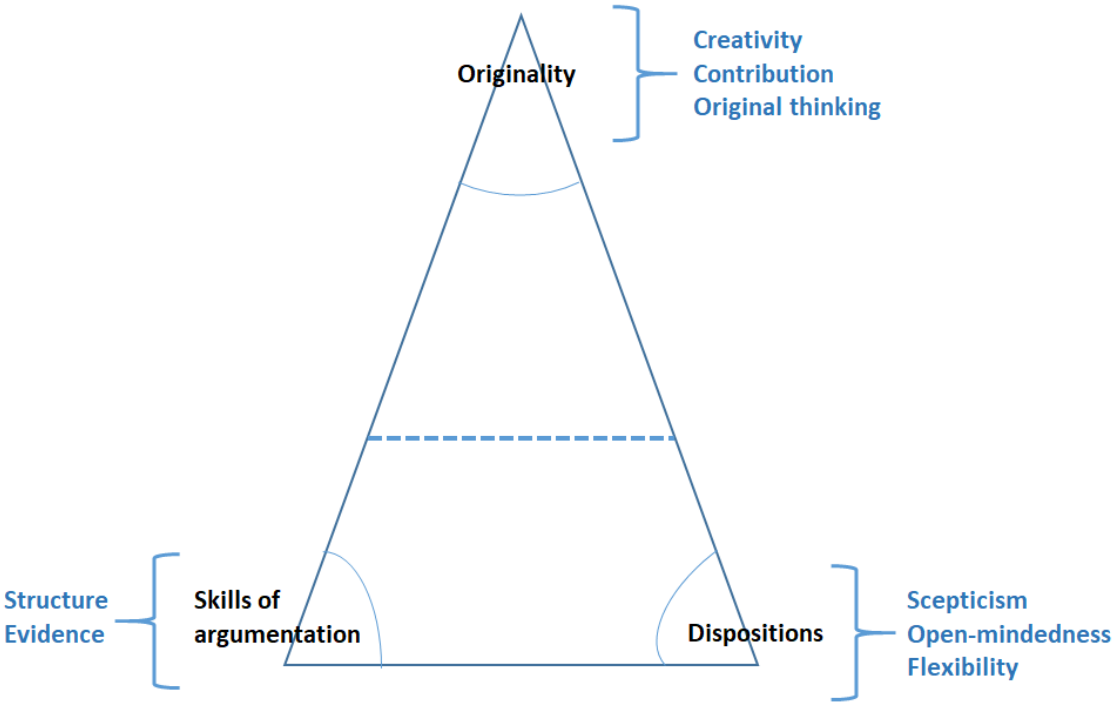

Fig. 5.1 The facets of undergraduate students as critical thinkers

The student as critical thinker has a developed skillset linked to argumentation. Structure and evidence are key to this. Students can demonstrate critical thinking through constructing valid, believable and logically structured arguments that avoid contradictions and which build from legitimate premises through to sensible and substantiated conclusions. What they also need is evidence in support of the conclusions. Lack of evidence detracts from the strength of the argument. At one level, evidence can take the form of referencing appropriate sources; at another, comparing

and contrasting sources; at another, critique of sources; at another still synthesising all of this into a coherent picture, as regards the literature, and as regards research findings viewed through the lens of the synthesised literature base.

The student as critical thinker exhibits certain dispositions. A propensity for scepticism marks out the student as critical thinker. The sceptical student doubts all, takes nothing for granted, questions and probes for answers. The open-minded student is curious and inquisitive, open to possibilities, able to see the bigger picture, willing to examine alternatives and weigh up the debates before adopting a justifiable position. The flexible student is not wedded to beliefs, is willing and able to change views where appropriate, is unfazed by problems and is able to adapt.

The student as critical thinker exhibits originality. This may be framed as creativity, exercised in various forms. The creative student demonstrates novel or innovative ways in which research can be set up and/or conducted. Originality may also be framed as contribution to knowledge, howsoever small. This may be theoretical, methodological, practical, professional or policy related, or a combination of these. Finally, originality may be framed in terms of students offering some original thoughts of their own, evidencing application of their intellect, their understandings, impressions, perspectives and positions. What would not constitute originality for participants are descriptive accounts and replicative-style approaches to research. Description is a well-worn pejorative term for work perceived to be rather basic, possibly superficial, certainly unoriginal, lacking in depth of thought and insight. Also cast as unoriginal are replication-style research studies. These offer a methodological blueprint to follow and a comparatively low risk route, but are subject to negative perceptions of "*dissertation(s) by numbers*" (*Florence*) and hence, unoriginality.

6. Theme 2 - Approaches to assessing undergraduate dissertations: convergence, difference and mitigation

6.1 Introduction

Chapter 6 is the second of two chapters in which I report my findings derived from thematic analysis of the data. Here, I develop Theme 2 – Approaches to assessing undergraduate dissertations: convergence, differences and mitigation.

Questions 2-4 of the interview schedule served to anchor the interview firmly in a situated context, orienting the early part of the discussion towards assessment in higher education and more specifically towards assessment of undergraduate dissertations. These were supported by Question 8, which targeted discussion at how participants have developed their knowledge and practices over time, and which naturally helped revisit areas covered by earlier questions.

Several points of convergence and difference in participants' approaches to assessing undergraduate dissertations became apparent from this analysis. I begin this chapter by examining two substantial aspects: firstly, holistic versus additive approaches to assessment; and secondly, expectations of originality, broken down into three further aspects: originality as contribution to knowledge, originality as fresh literature at the findings and analysis stage (evidencing original thinking), and originality as methodological creativity.

Also coming from the analysis are points concerning the institutional mechanisms installed in order to mitigate differences in approach and maintain assessment consistency, namely use of assessment criteria, support for and the exact application of which can and does vary, and internal moderation (double marking for undergraduate dissertations at the host business school). Comments relating to a third mechanism, external examination, were not substantial enough to warrant inclusion here.

6.2 Points of convergence and difference

6.2.1 Holistic versus additive approaches to assessment

A third of participants described their approach to assessing undergraduate dissertations as holistic. They see a dissertation as constructed from different parts, some of which may be good, others less so. Assessment is a matter of balancing these parts and assessing the piece as a whole, rather than allowing one good part or one poor part to colour the piece and define the result:

Florence: *Definitely more holistically….I'll be looking mainly at, actually all chapters are important, I don't think you can say there's any chapter in a dissertation that you can skim past, it all has to come together.*

A holistic approach fits with the assessment criteria, which essentially describe the overall required performance at and within several undergraduate grade bands:

Viola: *It's holistic. The mark is holistic. Otherwise, we would be in trouble wouldn't we? Because we have our criteria, we have to mark on all, right?*

Olivia too advocates a holistic approach, concerned that whilst there is always an element of an assessor starting to form a view about a dissertation as they read it, it is important to resist allocating a grade too early. Olivia is of the opinion that assessors should not put a grade on a dissertation until the end when they can form a holistic impression, suspending judgement until the dissertation is viewed in its entirety, conscious that quick and perhaps premature judgements are part of human nature:

Olivia: That's human nature. One thing that I probably am more aware of now is not to pigeonhole something absolutely until I've got to the end…I try and make sure that I don't consciously pigeonhole it too much because I can think of dissertations in the past that I've read which have had I wouldn't say scant literature review but one that could have been much better written, much better presented, but then have gone on and collected some highly original data, done something with really good methodology, and got some really interesting results and analysed them. And if I'd have been doing it in a too much an additive way I may have pigeonholed it too early. So I try and force myself to make some notes and then go backwards and sit and synthesise.

The difficulty lies at the end, in balancing the good and the less good areas of a dissertation to arrive at a holistic determination as represented by the overall awarded grade:

Olivia: *Sometimes that's where the problem is because you're like, well, actually it's really good here, but it falls down there, so where do you end up giving it a mark?*

Judging this balance is thought to be subjective and potentially arbitrary by some participants, four of whom consequently eschewed holistic assessment in favour of building additive grades:

Milo: *I think marking, in getting the idea that I go through it sequentially in the different sections, and if you look, when I look at my feedback that's how I'd mark it.*

Zhao: *I focus on basically three elements. The first is the theory, secondly the methodology and thirdly and very importantly the fit between the theory and the methodology. I break down the marks into these three areas, I would say, and perhaps theory 40%, methodology perhaps 20 to 30% and the rest for the fit.*

What Milo and Zhao are talking about is the construction of an overall grade bit by bit, for example, splitting the grade and allocating separate constituent grades for each chapter. This may help with standardisation, the assessor constructing a defensible grade that builds from separately assessed constituent parts, and which therefore accounts for any performance differences at each stage of the dissertation. From a practical perspective this may compare favourably with a holistic approach, which has its difficulties for the assessor who must mentally balance all aspects of a dissertation into one overall grade:

Evelyn: *I think the marking scheme…I think it really helped us because you were able to say, roughly a third for the literature discussion, a third for the method, bits of handling the result bits, wasn't it something like that and then a third for the conclusion and discussion and 10% for justifying the overall context of it and your professionalism…and because of that if you had a brilliant literature review where all the criticism was it compensated, and then not such a strong backend at least you got the credit at the front. Whereas I tend to think here because it is open it overly, if you've got a weak place, it overly depresses the marks in all of it…***(Note: Evelyn was referring here to a marking scheme used at a previous institution she had worked for)**

However, participants who advocate a holistic approach to assessment were set against this additive-based alternative:

Harry: *…when I say I write comments on each chapter I don't then award marks for each of these chapters and then add them up.*

Olivia: *I know, and I've seen that [an additive mark] before too. I think sometimes in a way that's "easier" – in finger quotes for your recorder - for an academic because you can hide behind that a little bit. Like, oh yes they did well here, they did, and then you're like well that's what the number adds up to so that's all I can give it, whereas I'm not sure that that's right. That's not how we think about the world and how we engage with the world, so no I'm more holistic.*

Of some concern with respect to the maintenance of consistency and comparability across assessors, two participants were evidently using this additive approach of their own volition and of their individual devising. Zhao, and now also Florence, seem to apply their own individually determined additive assessment schemes on a staged basis, notwithstanding the official criteria which as I have commented presuppose holistic assessment through describing how the overall performance of a piece should present at each of several percentage grade bands:

Florence: *What I did is I created my own marks sheet which basically breaks down the key sections, so intro, literature, method, results, analysis, discussion, conclusions, referencing, which I think is really important as well, and then a few of those little bits and pieces.*

6.2.2 Expectations of originality

6.2.2.1 Originality as contribution to knowledge

There was a definite split of opinion over whether or to what extent originality should be required of undergraduate students.

Of the eighteen participants who commented, a third of participants talked in terms of expecting some level or aspect of originality from undergraduate students:

Florence: *I know I've got colleagues in our department, in marketing, who expect to see an original contribution to knowledge from even an undergraduate dissertation, as well.*

Mason: *…contribution to knowledge, originality, you're expecting that. At PGT level and then to a lesser extent at undergraduate level, are they able to? Is what they're doing actually going to, yes I'm still looking at it advancing our knowledge in a particular area, even though it may be reasonably narrow and so forth.*

Some even interpret the assessment criteria as actually directing assessors to expect originality for the award of higher grades:

Theodore: *I think originality does make a difference because when you start to look at the top end of the grade descriptor it talks about such types of work are only rarely found, or whatever the exact phrasing is.*

Florence, however, would take issue with this, claiming that the module handbook, which includes the applicable assessment criteria, does not specify or imply that originality should be required at any percentage band:

Florence: *It's never explicitly outlined in the handbook that they're expected to make an original contribution to knowledge, or do something massively different.*

Florence: *It's like, yes, some colleagues when they're looking for an original contribution from an undergraduate dissertation, like ok, well that's interesting, but I don't think that's ever actually going to change my perspective unless it's written in the handbook and it changes that's what's written in the handbook of what's expected of them.*

The issue is one of academic levels. Originality obviously becomes more important the higher you go up the educational ladder. For example, that originality is required at doctoral level is a given and collectively understood:

Olivia: *And that's what required at a PhD level because you're supposed to be making an original contribution. You're supposed to, there should be something different about our understanding of the world that comes out of a PhD.*

Florence: *…they're pushing the whole the original contribution to knowledge. In my view it's PhDs where I would expect that's where at the level I would expect to see…*

However, when it comes to undergraduate students, two-thirds of participants were clear that assessors should not expect originality:

Reuben: *I don't think there's any need to be original in an undergraduate or a masters dissertation. I wouldn't even see that in a masters by research.*

There are clear differences of opinion therefore as to expectations of originality. Differences here can manifest in assessment outlooks and therefore grades. For example, *ceteris paribus* those who require originality but do not see it are likely to attach a lower grade to a dissertation than others who do not seek it or expect it in the first place. Differences on this issue have come to light for participants in disagreements at joint assessors meetings to discuss dissertation grades:

Florence: *And I find that, I contrast with and, they tend to be the most heated discussions when it comes to agreeing marks and they're pushing the whole the original contribution to knowledge.*

The feeling for some participants is that absent originality, students will struggle to demonstrate any real understanding and their research will lack those value-added aspects of application, synthesis, insight, etc. that are so important to demonstrating a contribution:

Thomas: ***…****must have some value added, must push something forward. There has to be certain amount of originality because otherwise it's really difficult I think to gauge whether they fully understood anything, but once they have to branch out a little bit, that's how you…*

Undergraduate students are, however, at a comparatively early stage in their learning. With empirical dissertations, for example, students can do competent research work at undergraduate level through collecting a fresh dataset and applying established methodology and methods to produce verifiable outputs (cf. replication-style studies, s.5.4.5). The vast majority of students are not going to make a breakthrough engaging in this type of research or indeed at that particular point in their learning and development. It would seem inequitable therefore to set the bar so high as to require originality:

Oliver: *…it would be rather unfair to require originality of thought in an undergraduate dissertation…*

Olivia: *… they [undergraduate students] really don't need to demonstrate a unique and significant contribution to knowledge…*

Harry: *I think the thing with, going back to the critical thinking thing, with a dissertation I think it's about what's the purpose of the dissertation? It isn't in itself to get published or anything like that. It isn't at the level of undergraduate or masters to make a contribution to knowledge.*

For these and other participants firmly in the 'originality is not required' camp, undergraduate students simply do not have sufficient time, or indeed the necessary full command of the literature in the first place, to branch out and produce something original that contributes to knowledge. This is not the preserve of an undergraduate dissertation and any expectations of contribution as originality would be unwarranted. There is, however, recognition that whilst not mandatory, originality in whatever form would inevitably stand out and most certainly impress with that aforementioned perception that this is the preserve of higher academic levels:

Arthur: *Well, the doctorate's the only one that's examined for a contribution to knowledge, so I'm not expecting them to discover something amazing, but sometimes they do and I've encouraged students in the past to publish. The one I gave 95 to got a publication out of it, but it doesn't have to.*

6.2.2.2 Originality as fresh literature at the findings and analysis stage

One practical question of judgement concerns exactly what literature students should bring into the discussion of their findings. Should this adhere rigidly to what is in the literature review or should the student now bring in fresh literature? The majority emphasis of participants is on the fit from research questions to the literature review to the methodology and methods through to findings and analysis, anticipating then clear alignment between chapters. For *Arthur*, however, whilst this approach would be indicative of a "*good student*", comparatively an "*excellent student*" at the findings and analysis stage would bring in additional literature thereby extending the overall literature base:

Arthur: *And then I'm looking for them going into the detail of the findings. A good student will relate it back to their literature review. A weak student just tells you the results. An excellent student will bring in some new literature at that point because they're thinking beyond where they started and they're trying to relate the findings to more things out there and then end by posing questions for future research as a result of that.*

Would this evidence some originality of thought? That in "*thinking beyond where they started*" and in "*posing questions for future research*" (*Arthur*) students are demonstrating some of their own original thinking? Two other participants, forming what would appear to be a minority view, lent further support to Arthur's strongly held view:

Florence: *…they've got to be able to reflect. It doesn't necessarily have to be with respect to the literature that they covered in the literature review, they can be bringing in new stuff, if necessary, but they have to really think about ok so how do my findings fit into the wider theoretical perspective, and why might my findings diverge?*

Oscar: *I'll say to students don't be afraid in your discussion to bring various new things in that might help you explain that, because in your literature your literature review will cover as much of the subject and maybe theoretical base that you can but until you've got your findings, and then you're looking at explaining them and discussing them, it might be that you're then bringing some new things to say ah this really explain this or helps explain this and so there might be a few little new bits that come in.*

There is sense in this. If students are critically analysing and thinking through their data then this could well open up new avenues and take them down routes not covered by the literature review, the scope of which is likely to be restricted due to its early staging and to achieve the necessary depth within a limited word count. Of course, what students may do is return to and augment the literature review to fit with the data analysis. This has the advantage of neatness but some would argue that this would not give a true picture of the research. In reality, it is unlikely that research is so well ordered that the literature review will encompass everything that is necessary to help explain the findings in advance of reaching them.

6.2.2.3 Originality as methodological creativity

Chapter 5, s.5.4.1 highlighted the suggestions that research could exhibit creativity through research questions but also the way in which these are pursued. Research could, for example, exhibit creativity and stand out from the crowd if the student has taken a risk. In practical terms, this could mean students branching out beyond the curriculum by doing something more methodologically complex than taught and anticipated:

Mason: *What it did was, it seems to me it gathered data together in a way that had simply not been done previously.* **[This came from the document elicitation part of the interview, in which Mason commented on a dissertation that achieved an 'outstanding' grade of 90% and the factors that determined that grade]**

Zhao: *And we also have a requirement for the methodology, for example, to be distinctive. They have to use regression if it's quantitative analysis. I have seen about 80 which uses, as here, structural equation modelling. We don't teach that in the department. Every year, very limited numbers of students managed to learn by themselves and they did quite well.*

Mason's example appeared to involve the student taking research methods from one disciplinary area and applying them to good effect to research questions situated in another disciplinary area in which such methods were not commonly employed. Mason deemed this very creative. However, Zhao's assertion that distinctiveness is required of the methodology is not backed by the assessment criteria of the host business school. Hence, this should not be seen as a pre-requisite for higher grades.

Now, whilst it is axiomatic that research should be robustly constituted and set up appropriately to answer the questions posed by that research, this does not necessarily point to creativity. However, the majority view, captured here from Arthur, is that what is important is that the methodology fits the study, not that it is creative *per se*:

Arthur: *I think they don't have to be innovative. That twenty-first questionnaire that I've read this year is fine if that's the right way to do it, and if that's the appropriate tool use it, don't try and be clever and screw a shelf on the wall with the blade of a kitchen knife instead of a screwdriver.*

This would seem to agree with a further comment from Oliver who was clear that technically competent dissertation research would be graded highly irrespective of the absence of creativity:

Oliver: *It won't stand out for me but I could still regard it as a piece of first class work given the parameters of accounting education. My predilection happens to be more for the thoughtful analysis rather than the technical but I try not to impose that on the students. I think if students can do an excellent technical dissertation then it's worthy of a first class mark.*

It seems that methodological creativity is not necessarily required then, but where it exists it is likely to be appreciated by participants who will reward it as evidence of originality and hence critical thinking:

Viola: *I would reward the creativity extremely high.*

Evelyn: *If I'd had a creativity mark I would have given them 85*.

Aidan: *…it depends on how well it's executed but then I certainly would, if somebody is doing something remotely different, not for the sake of being different but you can see that they are working towards something which has some promise, then I think the person should be given some reward for that…*

Continuing this theme, in Chapter 5, s.5.4.5, I highlighted the suggestion that research that adopts a replication-style approach is unlikely to exhibit originality and hence critical thinking. Participants did not dispute this. However, a small majority of respondents on this matter were positive of the approach, arguing that it provides a lower risk route for students to undertake research. A safer option then, albeit compromised, for busy, time-limited, assumedly academically under-developed undergraduates. It offers a path of less resistance and a relatively safe way of constructing research that will achieve at least a pass grade. It just makes sense therefore:

Milo: *Play it safe. Don't stick your head your head up above the parapet. You've only got a few months to do this. If it all goes pear-shaped the consequences can be dire. The upside, to put it in a finance thing, there's massive downside risk and the upside potential is very limited because they are not going to, unless they're outstanding or exceptional, there's no benefit to taking a…*

Olga: *So they go for some very well-researched topics, standard methodology, and maybe they choose different datasets from what has been in the literature or the same dataset but with the most recent data. On the one hand it's a sure bet that they will pass if they don't do anything outrageous and then on the other hand it's just plain boring.*

It seems that whilst the apparent safety first approach to methodology and methods of a replication-style study may not excite, and may be regarded as unoriginal and therefore less than critical, that it is “a sure bet that they will pass” (Olga) might appeal to undergraduate students. More than this there was a further suggestion that if this type of research is competently set up and executed then this can pay dividends with the grade beyond simply passing:

Milo: *And looking at some of the dissertation topics over the past number of years the idea of why is everyone doing mergers and acquisitions? They are playing it safe. It’s easy. They can replicate a study that’s gone before, change the dates, change the companies, and if they do it well they’ll get a 70, if they do not so well they’ll get a 60 odd.*

Evelyn: *…actually in some ways it’s easier to get a first if you’re pragmatic and pedestrian than it is if you try and do something that matters and that doesn’t seem to be right.*

Something of a ‘no-brainer’ then?

What students actually seek to investigate should be embodied within and evident from their research objectives/questions (again hereafter referred to simply as questions for brevity). Students need to identify and choose suitable methodologies and methods for framing and conducting their research in answer to those questions, and they need to be clear on exactly why their choices are fitting for their studies. This in itself requires some research into alternatives and potentially some critical thinking on the part of students to make the approach their own:

Fiona: *I do look to see whether they are able to identify an appropriate research method for the question…*

Zhao: *…in the methodology chapter, I want to see the students have some discussion because no methodology, no method is perfect, so they need to argue why I choose this rather than the other. So what's the benefits of doing so? What's the cost of not choosing another one?*

In practical terms, it does not or it should not matter if the student has taken a quantitative or qualitative approach. Rather it is important to establish that the chosen approach was apt for the study:

Florence: *I also really want to see that they've acknowledged why their chosen approach is the most relevant one. So not just to dive in and start describing a questionnaire but to actually say ok, given the research objectives and the hypotheses that have been outlined in the previous chapter, this is why a quantitative study is most important or most relevant, how it benefits over a qualitative study or vice versa obviously.*

Appreciably, however, this may be an issue and a source of difference for academics who have strong views on this and who would tend to favour certain methodologies and methods with which they are comfortable, which fit with their epistemologies and which fit with their views, often strongly held, of how research ought to be done over others:

Hanna: *I know that some people are always very picky on the methodology. You have to use X methodology, which I think is totally wrong.*

6.3 Mitigating mechanisms

6.3.1 Use of assessment criteria when assessing undergraduate dissertations

Half of participants considered that assessing for critical thinking necessarily relies on the professional judgement of the assessor:

Oliver: *It's a case of feel the quality basically, which isn't very scientific, but then I don't think necessarily that an assessment process should be cut and dried. I don't think it should be a tick box exercise. I think it should involve the assessor using judgement.*

Milo would suggest that as an academic this is something that he does intuitively, implicitly rather than explicitly. Similarly, over time academics develop "a feel for it" (Isaac). Professional judgement combines knowledge, understanding and experience with gut instinct, and plays an important role in assessing students' work. Assessment is not just a mechanical applicative process in which instinct is negated:

Milo: *The thing is that I would say that marking those, it is so I can, I wouldn't say it's a gut feeling because it is based on, I think it's based a sound marking criteria and stuff, well maybe not, but you know you get a gut feeling.*

However, for some participants assessment should not be left to professional judgement alone because this would invite inconsistency. Criterion-referenced assessment practices serve to mitigate this. The need for students to be 'critical' is felt to be integral to the assessment criteria for undergraduate dissertations, particularly as the grade scale progresses upwards:

Zhao: *So no doubt being critical is part of our standards.*

Olivia: *It would be the demonstration of the critical thinking that would really shift I think between a merit and a distinction, or a 2:1 and a first.*

There was majority consensus on the overall use of the assessment criteria. Participants generally agreed that these ought to be followed, albeit to a greater or lesser extent, when assessing dissertations:

Reuben: *I have it in front of me when I mark. I keep it in front of me. When I'm assigning a grade I'm thinking about what is the criteria…*

Florence: *I think it can be useful to go back and look at the grade criteria and remind myself, certainly at the start of marking process and then periodically going through about what those grade criteria actually are and again apply them to.*

This apparent willingness to comply is not simply because the institution by virtue of policy and process requires it but actually, and more positively, because participants see the value in doing so for themselves as assessors, it has utility for them. For example, the criteria serve as useful reminders of what is required to achieve different grades, particularly for those scripts straddling the border between grade bands:

Olivia: *Sometimes if I'm sitting there going oh is it this or is it that, then having a look the grading criteria reminds me ah yes it is that level or it isn't doing that, so it can only be this or whatever.*

Oscar: *If I was really struggling with one between let's say a class boundary, I would then go back to the grade [criteria] and think well I've got a good idea now where this is but that that would probably inform my overall opinion as to where the line might be drawn as to whether it just is a good 2:2 or whether it's just enough for a 2:1.*

Reuben, Florence, Olivia and Oscar all profess that they use the assessment criteria but appear to take different approaches to this. Reuben is constantly checking against the criteria as he assesses ("keep it in front of me"). Florence refreshes her memory of the criteria at the outset and then checks against the criteria periodically by way of reminders. Olivia looks to apply the criteria to help make a decision whenever she is unsure. Oscar similarly but only when deciding borderline cases that straddle grade boundaries.

A third of participants commented that whilst they used the criteria they did this only post hoc as a check on their already formed grade rather than as an integral assessment tool. Essentially, they first assess a dissertation without recourse to the criteria. They then subsequently refer to the criteria purely to help judge if the given grade is appropriate and credible. This then forms a valuable checking mechanism helping to achieve consistency across dissertations:

Isaac: *That's like seventeen consecutive years I've marked these things and you know what our standards are, you know what you are marking to, you know what the [criteria] are. So they're almost there as a second check. Have I done this right? Does it accord with the [criteria]?*

Mason: *I use it [assessment criteria] at the end if I'm honest…I'm aware that if anything I'm probably on the slightly conservative side, so I try to be sure that I'm not being unduly harsh on the students. So it's at the end where I wanted to make that overall summary that I then go back and double if, sometimes I don't because I just know this is clearly just scraping a pass, is clearly 52 so what's, is it fair or satisfactory or something, what's the key word for that? But I will usually reread that grade criteria at that point before I make my overall judgement to say yeah that is a good description of this piece of work, so therefore it is 68, therefore it's very good.*

Participants who discussed this generally shared this sense of value in applying the criteria at the end as a check. However, for Milo it seemed more a case of simply going through the motions in retrospectively applying the criteria:

Milo: *The grading, the things, they're not fit for purpose, I don't think, but we have to fit into them. We kind of mainly now make it up after the fact.*

Milo: *I fit the grading descriptors afterward. It's fitting that into the grading descriptors. I have my own set of what is good and what not and then the grading descriptors are an afterthought, if you like.*

Milo seemed to lack faith in the criteria. Nonetheless, he does claim to apply the criteria, albeit as a final sense check on his awarded grades.

Evelyn and Mason expressed concerns over alleged non-use of the assessment criteria by some colleagues:

Evelyn: *I honestly, don't you dare tell anybody this, have a suspicion that some people don't even read them at all.*

Mason: *I actually have very little idea [referring to how colleagues use the criteria]. I have some evidence, for example, from this latest bunch of marking of people apparently not using them at all…*

Five participants seemed to refer to having adopted their own adjusted or alternative assessment criteria, for example:

Milo: *I have a system for myself of what I'm looking for because that allows me to be consistent.*

Arthur: *I'll be looking for particular things in terms of the marking criteria, my own, I suppose I've got this unwritten list in my head of things I look for in a good dissertation.*

Three participants acknowledged that they did not specifically consult the criteria when assessing dissertations. Oliver and Harry were by their own admission conscious of the criteria but not reliant on applying the criteria first hand, being confident in their knowledge and understanding of the criteria from past assessing experience:

Oliver: *I've obviously already done a host of undergraduate exam marking so the descriptors, the grade [criteria], are in my mind anyway informally. So I've got that sort of background but I don't particularly refer to any…while assessing dissertations.*

Harry: *I'm not saying that when I'm marking an individual piece I've got the grading criteria there. I'm pretty reasonably confident in, I know 60 to 65 is good, then I know the other's very good, and then excellent at 70s, outstanding in the 80s.*

Harry's comment is in itself interesting as 'excellent' forms part of the description of a grade of 70+ per the criteria but strictly speaking 'outstanding' comes in from 76 not 80+. This comment then is broadly in line but not fully in accordance with the criteria. Harry is seemingly confident in his grasp of the criteria, now apparently internalised from long-standing use, but perhaps is not quite accurate or up-to-date with any changes made?

The third participant, Olga, appeared more reliant on a sense of where the grades lie based not on application of the criteria but on an iterative scaling process involving relative comparisons across scripts applying standards observed from assessing dissertations in previous years. This suggests that Olga's approach is more norm-referenced than criterion-referenced:

Olga: *Not very strictly [referring to use of the criteria]. I rely more on my experience from the previous cohorts of students. I will probably think first more broadly in terms of whether it's a first class or whether it's distinction, say. First I will place it into a class and then I will think about maybe is it closer to the bottom or the middle or the top of the class.*

Collectively, these excerpts would appear to highlight a general commitment to the criteria for the most part but not an altogether consistent approach to application. Yet, with respect to the students, the widespread use of the criteria by assessors is thought crucial in presenting the necessary confidence that assessment is equitable:

Theodore: *Well, we have to stick to our criteria so that we are fair to all students.*

This guards against perceptions of a 'post code lottery' effect, whereby the grade a student receives for his or her dissertation is subject to the whim of the assessor concerned and hence the luck of the draw.

A third of participants discussed the importance of the criteria in maintaining assessment consistency. Views on this were twofold. Firstly, that the assessment criteria helped with self-consistency in assessing across dissertations:

Isaac: *So if I've got a batch of twelve I want to make sure that I've marked the twelfth one the same way that I've marked the first one. So you use them [assessment criteria]…*

Fiona: *So after I'd looked at a couple of them [undergraduate dissertations] I did a check against the [assessment criteria]. I did look at them again further down after I had done a few more just to make sure that I was being consistent with them as I went through them.*

Secondly, that this applies equally to consistency across assessors with the assessment criteria playing an important role in facilitating this:

Florence: *…well those are the [assessment criteria] that we're given, so I think if we're going to have consistency across markers, that's why I've always got to go back and refer to, I think.*

Mason: *I would say the majority of the time I'm within the typical a few marks outside of other markers and therefore I've got to assume they are also using them.*

Zhao: *Basically, we have to stick to the criteria, we have to contrast the work against the criteria. That's how we can achieve agreement.*

Florence, Mason and Zhao, as advocates of consistency, recognise the importance of each assessor approaching the task informed by the criteria. For Oscar, this would include assessing for critical thinking:

Oscar: *I think it is important as well that your view of critical thinking, or your view when you're looking down on marks, it is informed by the grade [criteria] and you can't just sort of say well this is my little view and that's the way it is, because then rather you would be criticised for saying well there's just one right answer and that's yours, but actually there is a bit more of a guidance beyond that…*

On questions of readability, a quarter of participants commented in favourable terms on the level of clarity afforded by the language used in the criteria and their worth to the assessment process as a result:

Reuben: *I think what happens is experience does help clearly but I don't think that you can go far wrong on the marking guidelines. They're pretty clear.*

However, if a key aim of the criteria is to achieve consistency across assessors then a potential problem here is the apparent level of subjectivity in participants' interpretations and application of the criteria:

Isaac: *The criteria themselves are a little bit open to interpretation…*

Penelope: *I think it [assessment criteria] is quite vague and I always, yeah, it's like I don't like the terms, it's satisfactory and fair isn't it? And I always think well what does that mean actually?*

It seems that some participants, whether dismissive of the official assessment criteria or not, are choosing instead to judge students' work against criteria of their own devising, internalised from experience and so individualised criteria which may or may not incorporate or compare with the official criteria, and which may nor may not be suitable for undergraduate dissertations. That said it seems unlikely that individualised criteria will deviate much given the likelihood that these will have been informed by the official criteria at some point, together with the calibration effects of colleagues' expectations within local assessment communities. Equally, however, it would seem important for consistency that assessors follow the guidance offered by the published criteria:

Thomas: *It's important I think to have a set of guidelines as to what are the classifications, what is expected of each level, rather than making it up for yourself.*

Assessors will have their own understanding of critical thinking and of how that might inform the assessment process. Appreciably, however, for consistency that understanding should tie into the official criteria:

Thomas: *I think they [assessment criteria] are pretty critical because they set for me an expectation of what I'm marking within. I could come up with my own expectation and deliver my own version of it but that wouldn't be fair to the students that get me because I mark harder or softer. So I think it's very important to have…extremely important, and then I'm trying to work within those.*

Absent the criteria, and/or substitution by individuals of internalised or tacit criteria, then personal preferences and enthusiasm or lack thereof for disciplinary and/or methodological approaches taken by students could override assessors' objectivity:

Viola: *…if we don't have those [assessment criteria] then the subjectivity becomes even more of a problem…Otherwise, we might be biased. Me because of the research question, might be other people because of the methodology, you know?*

More credit, perhaps unduly, might be given to research topics and questions or the adoption of methodologies that are favoured by the assessor. Correspondingly less credit may be given for topics, questions or methodologies of limited interest, or for which the assessor has little or no experience or appetite. That would seem unfair to students who may choose topics borne of their own interests and who often make pragmatic methodological choices on what can be done within the limitations of time and available resources. Consequently, the criteria set general expectations applicable across the board, seeking to move assessors away from personal biases and towards neutrality:

Aidan: *But then I think, if everybody was just to train and to be quite neutral, some people are more passionate about certain topics than others and then this is where the weighting comes in, or they have been to an institution where there is a particular emphasis in certain kinds of weighting, right? But I think a lot of academics are surprisingly, you know I think if you just get people to think and work as neutrally as possible you won't get marks that are too far off, you will actually have quite a harmonised set of marks.*

6.3.2 Internal moderation (via double marking)

The host business school has a varied academic base:

Hanna: *…we have we got a lot of new colleagues that come from all kinds of universities with all kinds of expectations.*

Conceivably, it follows that conceptions of what is required or valued in dissertations, more specifically what critical thinking is and how it manifests in dissertations, can and does vary.

Grading variation is particularly observable with newer academic members:

Theodore: *If you mark with some colleagues who are new, or new to the UK system, then there might be quite a large divergence.*

Engagement with colleagues in local assessment communities is hugely important to developing shared views of critical thinking. Academics can develop through challenging themselves, not just in their own research but also by engaging with colleagues, particularly experienced ones, having intelligent discussions, perhaps even heated debates, whether maintaining an argument of contested points or shifting their thinking. Such interactions and experiences are all important for developing an understanding of critical thinking:

Fiona: *I think just hearing people with very good brains talk about different situations… conversations with academics with really good brains on particular different issues…*

Olivia: *Obviously, some people can challenge themselves through their reading and so on but I think that actually engaging with your colleagues and having discussions around things and being challenged about things and having to make an argument or shift your thinking or, those all are important for developing your critical understanding, your critical abilities as well, so that you can understand a perspective from someone else's viewpoint. You might not agree with them but you can see, you develop ways of integrating that into your knowledge. And I think that's very difficult to do in a siloed position.*

When it comes to the practice of assessment, newer academics learn from more experienced colleagues, evolving their understandings and moderating their expectations, becoming more confident as assessors. This is part of the fabric of academic life. Experts and novices meet, they discuss and debate, and novices grow as a result:

Mason: *…from the very early days when you're a fairly green supervisor and you're not quite sure of stuff and you go through that first batch or two and you second mark with people and you're oh right ok, you begin to get a sense of where you should be. That's really helpful in those early days.*

Florence: *…it evolves, it very much evolves over the years. I think within the first three or four years I came to a very, a much more confident place in terms of supervision and in terms of having a bit of authority when it comes to actually discussing, working out a mark and then discussing and arguing it with a colleague.*

Given that the academic base at the host business school is so varied, and given that academics can learn from each other, then institutionally there should be mechanisms for sharing knowledge to help academics to standardise when it comes to critical thinking and assessment. A key mechanism in this context is the double marking process, where two academics each assess a dissertation separately and without recourse to each other and then meet to discuss their respective grades and the rationale for these with the aim of agreeing a final grade. Evidently, questions of whether and to what extent students are 'critical' do arise when assessors are comparing notes and discussing first and second grades on dissertations:

Theodore: *I would hope that when agreeing dissertation marks with colleagues, we'd often talk in terms of how critical students have been…*

Oliver: *…talking to colleagues, I came to understand that perhaps my expectations were unreasonably high and that I had to acknowledge the fact that I was dealing with young people just embarking on their professional education or their academic education and I had to tailor my expectations accordingly.*

Two-thirds of participants referred specifically to the importance of the double marking process, appreciative of what they can and do learn from it. Meetings bring out comparative views, helping assessors to get a sense of what they are each looking for, to understand what they are picking out as critical or uncritical in dissertations, to get a feel for where they and where their expectations are relative to others:

Olivia: *And also your own marking but the second marking practices and things like that so you can see how other people look at the same piece of work and what they see in it, and I think that that's really important and getting that thing too.*

Aidan: *I think maybe joint dissertation marking, just understanding people's criteria and then getting them to rationalise, so I think I think in a sense like double marking for something that is more subjective is important.*

For Theodore, a process of calibration does occur through the internal moderation process:

Theodore: *…but no-one had really talked to you about applying these [assessment criteria] and the various systems. So I think as an early marker it was very much a process of calibration. In the first couple of years you calibrated your expectations of work to particular standards by marking with a range of different people. It was a process of calibration, unquestionably.*

It follows that the longer academics are part of the host business school, and the more they interact in this context, the closer their views will align on what is and is not critical in dissertations, and the closer their respective grades will be. This it seems is subject to the continuing stability of the academic faculty, as practices and understandings infuse through departments:

Evelyn: *I'll share what I used to do and used to have. So, I think how I thought about it is quite clearly infecting us as a department as we go along. There is sort of a virus of this going on.*

Calibration appears inevitable when you work with others over a period. However, calibration takes time to take effect and is seemingly reliant upon experienced colleagues sharing their views:

Theodore: *I think it depends on colleagues. I find that if I mark with colleagues who are experienced it's very often the case that we're one or two marks apart.*

Isaac: *If you mark with somebody who is quite experienced you find that your marks aren't that far apart. So you are looking for the same things. But you do find, and this is coming back to experience, you do find that there is often much more divergence, you definitely see it in dissertations because you are blind double marking. So there is no anchoring to the other person's marks. And over the last couple of years I marked with* ***[redacted]*** *and we were never more than two or three marks apart. Yet I've marked with other people and you're 10, 15 on dissertations. They were less experienced.*

Arthur: *The interesting thing as I've got older is that I find now, when I sit down to agree second marks with people, whether I'm the first or the second marker, I'm much closer to other people than I used to be. I used to often be miles away when I first started with somebody more experienced as the other marker, now I'm much closer.*

6.4 Theme 2: Summary

Two specific areas of interest where points of convergence and difference were apparent came out strongly from the discussions.

The first is holistic versus additive approaches to assessing undergraduate dissertations. Some participants described their approach to assessment as holistic, meaning they read a piece of work as a whole and mentally balance its constituent parts into one overall grade. Others construct a grade on an additive basis, for example chapter by chapter, leading to a composite grade. This inconsistency is of concern, more so taking into account the apparently individualised nature of additive approaches with no official additive assessment structure to follow.

The second is expectations of originality. I have broken this down into three aspects. Firstly, opinion was divided on whether undergraduate students should be expected to contribute to knowledge. Secondly, participants diverged on the issue of whether or not to bring in fresh literature at the findings and analysis stage of a dissertation potentially evidencing original thinking. Thirdly, methodological creativity would point to originality, at least for some participants; and whilst participants generally held the view that a replication-style approach to research is unoriginal, they differed in their support for this approach.

The problems presented by differences in assessment approaches are subject to mitigation by established mechanisms, namely criterion referenced practices involving the use of assessment criteria, and internal moderation processes (double marking for undergraduate dissertations at the host business school). However, the evident inconsistencies in approach could undermine these efforts at standardisation and hence work against achieving consistency in assessment.

Most participants agreed with the application of assessment criteria when assessing undergraduate dissertations, as mandated by the host business school, noting that this is important for consistency. Actual practices varied, however, from applying the criteria throughout the assessment process, to applying only after the event as a post hoc check, to applying only intermittently as a review mechanism or applying only to help decide borderline cases sitting on grade boundaries. Essentially the assessment criteria are generally supported but inconsistently applied. Most but not all participants would have regard to the official criteria, either current or internalised from long-standing use, with some apparently substituting their own criteria. This too would work against consistency.

Engagement with colleagues is hugely important to developing shared views. Internal moderation via double marking of undergraduate dissertations facilitates interactions between assessors. Questions of whether and to what extent students are ‘critical’ do arise in such interactions. Assessors can work through these and other respective differences to determine final grades. Subject to the stability of the academic faculty, and given enough time, this mechanism can help to develop shared understandings and calibrate assessment within local assessment communities.

7. Discussion

7.1 Introduction

Here, in Chapter 7, I discuss the findings from Chapters 5 and 6 set against the three research questions listed in Chapter 1, RQ1-RQ3 inclusive, taken in turn and examined in relation to the literature base from Chapters 2 and 3.

The findings in Chapter 5 cover the ways in which participants conceive of critical thinking in relation to undergraduate students, more specifically final year undergraduate students given the dissertation context in which interview discussions were located. These findings feed into RQ1 and RQ2, enabling me to build a picture of what critical thinking is for participants, i.e. a composite of skills, dispositions and originality, to which I add knowledge (with reference to RQ1), and of how that concept of critical thinking is operationalised through assessment of undergraduate dissertations (with reference to RQ2).

The findings in Chapter 6 sit in response to RQ3, covering the challenges for consistency in assessing for critical thinking in undergraduate students' dissertations. Drawing on how participants conceptualise critical thinking (RQ1) and how they operationalise this through assessment (RQ2), I examine now points of convergence, differences and mitigation split into three parts: firstly, differing expectations of originality; secondly, holistic versus additive assessment approaches; and thirdly, sharing understandings and calibrating assessment.

I conclude this chapter with a brief summary.

7.2 RQ1: How do business school academics conceptualise critical thinking?

7.2.1 Cognitive elements - skills of argumentation

Critical thinking in higher education is fundamentally about making reliable and persuasive judgements substantiated through sound assessment of evidence from various sources (Moon, 2008). Skills of argumentation are central to this. Accordingly, these provided the focus of much of the discussions at interview and the core of participants' conceptualisations of critical thinking. This mirrors the first part of the Delphi Project consensus statement (See Facione, 1990) and follows the argumentation as critical thinking approach of the study skills texts prevalent in higher education (see, for example, Cottrell, 2017).

Participants further discussed what for them would make for good arguments demonstrating critical thinking. In the findings, I have grouped relevant points under 'structure' and 'evidence'. These subheadings echo key areas of focus exhibited by standardised tests for critical thinking. The HEIghten™ critical thinking assessment tool, for example, centres on analysing and evaluating the structural integrity and evidential credence of arguments, premised on strong arguments being structurally robust and evidentially sound (Liu et al., 2016). I suggest that the makeup of such tests reflects the views of the academics who construct them. This understanding of critical thinking as manifest in argumentation is common in higher education (see, for example, Andrews, 1995; Davies, 2013, 2015; Mitchell, 1994; Mitchell & Riddle, 2000; Scott, 2000; Wingate, 2012). The aforementioned critical thinking self-help guides are premised on this widely held construction of critical thinking as skills of argumentation. Evidently, the business school based participants in my study share such views, echoing the skills-based view of critical thinking, which builds on logical reasoning and is characterised in terms of cognitive and interpretive skills and mental processes connected with deconstructing arguments, and assessing and evaluating existing scholarship (see, for example, Bailin et al., 1999a; Carrington et al., 2011; Davies, 2015; Ennis, 1987; Espey, 2018; Fisher & Scriven, 1997; Halpern, 1996; Harrell, 2011; Kuhn, 1991; Liu et al., 2016; Moon, 2008).

7.2.2 Dispositions

The Delphi project consensus statement (see Facione, 1990) covers cognitive skills, connected with argumentation and judgement, and dispositions. Similarly, the critical thinking in higher education models of Davies (2015) and Halonen (1995), for example, take the cognitive thinking skills and add attitudes/dispositions as important elements of propensity, meaning "an inclination or tendency to behave in a certain way", which serve to motivate the exercise of the cognitive skills elements (Davies, 2015, p.55). A person needs the skills but also the propensity to exercise them to engage in critical thought (Butler & Halpern, 2020; Halpern, 2014; Halpern & Sternberg, 2020).

The literature consolidates a list of dispositions, expanding upon the Delphi Project, *inter alia* open-minded, fair-minded, curious, inquisitive, flexible, reasonable, orderly, diligent, persistent, objective, well-informed, honest, prudent, willing to consider and ready to self-correct (see, for example, Bailin et al, 1999b; Calma & Davies, 2020; Ennis, 1985b; Facione, 1990; Facione et al., 1995; Halpern, 1998; Halpern & Sternberg, 2020; Paul, 1992). The dispositions that I have interpreted from the data as amalgamated bear comparison with the Delphi Project consensus statement (refer s.2.2.1) and the detailed list of dispositions specified in Davies (2015), i.e.:

<table>
<tr><th>Dispositions arising in relation to self</th><th>Dispositions arising in relation to others</th><th>Dispositions arising in relation to world</th><th>Other</th></tr>
<tr><td rowspan="2">Desire to be well-informed</td><td rowspan="2">Respect for alternative viewpoints</td><td>Interest</td><td>Mindfulness</td></tr>
<tr><td>Inquisitiveness</td><td>Critical</td></tr>
<tr><td>Willingness to seek or be guided by reason</td><td>Open-mindedness</td><td rowspan="9">Seeing both sides of an issue</td><td rowspan="9"></td></tr>
<tr><td>Tentativeness</td><td>Fair-mindedness</td></tr>
<tr><td>Tolerance of ambiguity</td><td>Appreciation of individual differences</td></tr>
<tr><td>Intellectual humility</td><td rowspan="6">Skepticism</td></tr>
<tr><td>Intellectual courage</td></tr>
<tr><td>Integrity</td></tr>
<tr><td>Empathy</td></tr>
<tr><td>Perseverance</td></tr>
<tr><td>Holding ethical standards</td></tr>
</table>

Table 7.1 Critical thinking dispositions (extract from Davies, 2015, p.58)

My list – comprised of scepticism, open-mindedness and flexibility - is on the face of it much less comprehensive. Essentially, I have interpreted a distilled set of dispositions from discussions with participants, i.e. scepticism (incorporating curiosity and inquisitiveness); open-mindedness (incorporating prudence, tolerance of ambiguity, appreciation of individual differences); and flexibility (incorporating persistence and self-correction). Mine is a summarised list, amalgamated from smaller parts into sections carrying more evidential weight. This is not to limit the dispositions that characterise critical thinking in higher education. Rather this serves to highlight those that are apparently of particular relevance to the participants in this study as business school academics.

7.2.2.1 Scepticism

Critical thinking is underpinned by a propensity for scepticism towards authorities and towards received wisdom (Moore, 2013; Rear, 2019). Students who are sceptical have an inquisitive attitude (Lai, 2012). They do not just take things for granted or as given. They do not passively absorb knowledge from experts and authority figures (Moon, 2008). They have the habit of mind to doubt, to question, to challenge. They have a desire to be well informed. They ask questions, probe, strive for meanings and justifications and consider alternatives (Barnet & Bedau, 2010; Carrithers & Bean, 2008; Perkins et al., 1994; Walker & Finney, 1999).

Invariably, participants viewed scepticism as a positive. Students might “develop a cynical attitude” (Aiden). Cynicism - listed as a synonym for scepticism (Collins English Thesaurus Online, 2020) – has negative overtones but the scepticism considered here is evidently ‘healthy’ not jaundiced (‘healthy’ with reference to Macpherson & Owen, 2010). With this mind-set comes a reticence to accept current orthodoxy and a willingness to challenge it (James et al., 2010; Jones, 2009; Mingers, 2000).

7.2.2.2 Open-mindedness

The critical thinker is predisposed to being open-minded. This also came through strongly in discussions with participants. Students as critical thinkers do not simply accumulate knowledge, but must work to overcome resistance and uncover that which is hidden (James et al., 2010). Necessarily then, students who are open-minded are also curious and inquisitive (Lai, 2012). They are appreciative of the fact that they cannot know everything, whilst being open to different perspectives and possibilities.

Thinking that is critical is applied with deep engagement to complex matters that are likely subject to different viewpoints (Moon, 2008). Analysis and interpretation of research data may well lead to multiple, alternative, even conflicting explanations. Students may, for example, have tested a theory and generally found that it holds to an extent but not exactly or fully. There might be areas where it does not quite work or where the theory is unable to explain the results adequately. Perhaps it is the wrong theory for the situation under investigation. Perhaps it is the right theory but parts of it remain underdeveloped. Perhaps there are certain limitations to how or where the theory operates. Participants expect students who are thinking critically to be open and alert to all such possibilities. The critical thinker is not defensive and is open to possibilities (Brown & Rutter, 2008). Students who are open-minded will in the midst of this uncertainty question all and actively look for other perspectives to help inform their understandings and explanations and to necessarily support their choices. They can, at least for a third of participants, also "*see the big picture*" (*Fiona*), i.e. they can see beyond the immediacy of what is in front of them through to greater concerns.

7.2.2.3 Flexibility

Flexibility goes hand in hand with self-correction. Students can demonstrate their flexibility by, having exercised their scepticism and open-mindedness, having persistently looked for, probed and sifted possibilities before taking positions, they then have the presence of mind to self-correct, to change their views and reform their positions in light of fresh perspectives and new evidence.

7.2.3 Originality

Originality in dissertations is said to manifest in various guises, subject it seems to disciplinary norms (Guetzkow, Lamont & Mallard, 2004; Lamont, 2009; Lovitts, 2007). Guetzkow et al. (2004, p.190) talk of originality in the natural sciences as "the production of new findings and new theories", whereas for social sciences (within which I would situate business disciplines), originality more broadly encompasses "using a new approach, theory, method, or data; studying a new topic, doing research in an understudied area; or producing new findings". This description is suggestive of creativity. Furthermore, Clarke and Lunt (2014) cite intellectual contribution as the mark of originality in the social sciences. I discuss creativity and contribution in turn.

That thinking critically involves thinking creatively was discussed by twelve of twenty-one participants. This mirrors the literature, which describes thinking critically as involving thinking creatively (Bailin et al., 1996b; Halpern, 2014; Moon, 2008; Paul & Elder, 2006). Creativity embodies originality (Brodin, 2016; Dumitru, 2019; Jackson & Shaw, 2006; Lovitts, 2008; Winter, Griffiths & Green, 2000). Originality is in turn closely associated with critical thinking (Bailin et al., 1996b; Halpern, 2014; Holbrook, Bourke, Fairbairn & Lovat, 2007; Moon, 2008; Simpkins, 1987). Certainly being creative in an academic research context can evidence critical thinking for participants. Being creative can involve exercising ingenuity and innovation in doing something that is novel and adaptive (Amabile, 1996; Anastasiadou & Dimitriadou, 2011; Bennich-Bjorkman, 1997; Brodin, 2016; Sternberg & Lubart, 1999). Creativity, for participants, relates to students thinking differently about their research topics and the problems at hand, evidenced perhaps by going beyond the norm in their approaches to research or in their thinking.

Beghetto (2013) and Bennich-Bjorkman (1997) would differentiate creativity from originality on the basis that creativity need only be novel whereas originality must be novel but also relevant or applicable. Beghetto (2013, p.15) speaks of students' original contributions being "task-appropriate and thereby creative". Pope (2005, p.xvi) similarly refers to creativity as providing something not only "fresh" but also "valuable". These would suggest that there has to be a point to the research, a worth, a usefulness even (see also Halpern, 2014; Halpern & Sternberg, 2020). This brings us to contribution. The student as critical thinker engaged in research, even at undergraduate level, is anticipated to make a contribution. The majority of the participants who commented upon originality were of this opinion. Some benefit has to be derived. There has to be a point to it and an outcome that adds in some way. Moore (2013), for example, highlighted the importance, as far as participating academics in his study were concerned, of students actually going beyond challenging the ideas of others to demonstrate some originality of thought in producing their own ideas and of making contributions to knowledge, howsoever modest.

7.2.4 Knowledge

Knowledge in this context encompasses subject knowledge and epistemological development. Critical thinking requires a deep knowledge of the subject (Halpern & Sternberg, 2020). Participants confirmed this, some indicating the importance of reading around the subject and having the base of subject knowledge as a prerequisite for thinking critically about it. How knowledge is perceived is also important to thinking critically (see, for example, Battersby, 2018; King & Kitchener, 2002; McPeck, 1981; Siegel, 1985).

7.2.4.1 Epistemological development

The given descriptions of skills and dispositions help describe the student who is disposed to critical thinking and whose perspective on knowledge is developed sufficiently to engage with critical thinking. Self-invested in seeing the world in a more questioning but also more personal way; constructive self-criticism and self-regulation; being alive to possibilities; sifting multiple perspectives with openness, fair-mindedness, prudence and discipline; having the flexibility to deal with problems and to change perspectives and positions. These aspects all resonate with the epistemological development depicted in the models of Perry (1970) and Baxter Magolda (1992). They are clearly not indicative of the lower scales, essentially dualist positions on knowledge at which students hold absolute interpretations of the world, where things are a binary choice of, for example, right or wrong, and where 'truth' can be known with certainty (Carroll, 2007). Rather, they are indicative of Baxter Magolda's (1992) third domain of 'independent knowing' upwards, similar to Perry's (1970) relativism. These presuppose developed thinking capabilities and relativism on the part of final year undergraduate students.

Where students reach the stage of becoming relativistic in their outlook they move towards forming and justifying their own personal perspectives judged in context and on the evidence in the midst of uncertainty, opposing viewpoints and alternative frames of reference (Baxter Magolda, 2002; Carroll, 2007; Moon, 2008). They recognise that not all views are equally valid, that some views are better than others are, and they are able to evaluate and judge these comparatively (Hofer, 2001). In taking justifiable positions, students will be showing that they view knowledge as constructed and understood in context (Baxter Magolda, 2002). Having analytically sifted multiple perspectives they construct their own perspectives and adopt a position judged in context and on the

evidence (Baxter Magolda, 2002; Carroll, 2007; Evans et al., 2010; Hofer & Pintrich, 1997; Moon, 2008; Ostorga, 2006; West, 2004). They think it all through, using the current bundle of what is known to figure things out for themselves. This is what participants expect of final year undergraduate dissertation students. The critical thinking student, exercising healthy scepticism, is operating as Brownlee (2004) suggests at these higher levels of epistemological development, engaged in a thoroughly disciplined, analytical and critical consideration of issues from all sides and with regard to all aspects and perspectives both expert and personal.

Participants further value originality of thought. They are interested in what students have to say about their own perspectives, having weighed up the information and evidence. This too fits with the higher relativist positions of the models of Perry (1970) and Baxter Magolda (1992). Students who demonstrate originality of thought do not reside at the absolutist end of the scale. They do not regard truth as known and unchallengeable (Brownlee, 2004; Carroll, 2007; Hofer, 2001). They recognise that experts do not have all the answers (Hofer & Pintrich, 1997; Moon, 2008). Rather, they demonstrate evaluative recognition of multiplicity, of differing views, of uncertainties and conflicts (Brownlee, 2004; Carroll, 2007; Hofer, 2001). They think for themselves, now generating their own perspectives and holding these as equally valid (Baxter Magolda, 1992; Hofer & Pintrich, 1997). This signifies that students are not just working to a dissertation formula, for example, are not just regurgitating the ideas of others but are taking ownership of the material, reconstituting it in their own thoughts, forming and articulating their own perspectives.

These models of epistemological development propose that students progress through development stages in how they perceive knowledge. In turn, students' epistemological development and their capacities for critical thinking are thought to be related (Moon, 2008). Appreciably, however, this is unlikely to influence business school academics, not directly at least. In truth, few if any business school academics will possess a detailed awareness of epistemological development, more specifically the models of Perry (1970) and Baxter Magolda (1992), much less have their views informed by these. My position here is not that these models influence business school academics' perceptions of critical thinking as conceived and operationalised through assessment. That is unlikely to hold. Rather, I am suggesting that these models mirror what participants want and expect from final year undergraduate students in relation to exercising and evidencing their critical thinking. Students are enculturated into academic communities in which

academics set the tone and ideally communicate their expectations to students, directly or indirectly, which students absorb. It follows that in capturing the epistemological development of university students, the models of Perry (1970) and Baxter Magolda (1992) are also arguably reflective of academics' expectations of students at different stages of their epistemological development. Students in the more developed stages are more disposed to engage in critical thinking and more successful in both doing so and in evidencing this, in line with participants' expectations.

7.3 RQ2: How are business school academics' conceptualisations of critical thinking operationalised through assessment practices?

7.3.1 Skills of argumentation

Participants regard effective arguments as being logically structured. Critical thinking as argumentation is often connected, as here, with the application of the formal structures of logic and reasoning (Bailin & Siegel, 2002; Erikson & Erikson, 2018; Moon, 2008; Pithers & Soden, 2000; Swanwick et al., 2014). Several authors have defined critical thinking as the ability to interrogate and pick apart an argument, comprehending the reasoning behind it, correctly assessing its statements, grasping inferential connections made, tracking inconsistencies, detecting fallacious reasoning, etc., i.e. processes of logical reasoning (see, for example, Browne & Freeman, 2000; Ennis, 1962, 1993; Halpern & Sternberg, 2020; Kurfiss, 1988; Lundquist, 1999; Pascarella & Terenzini, 1991; Tsui, 2002).

Critical thinking is not necessarily defined by logic and reasoning, but with argument being such an important skill component of critical thinking, and logic and reasoning constituting important components of argument, these are necessarily important tools for the student critical thinker. Logic and reasoning provide the spine of an argument and the glue that holds it together. Students who are thinking critically will construct arguments that build logically from legitimate premises through to valid conclusions. If the logic of the argument is clear and cogent then this will go some way to satisfy participants as to the critical thinking displayed. Whether the argument is actually 'correct' appears moot. An argument can be entirely logical as constructed but this does not necessarily mean that it is correct (de Bono, 1983). Checking the correctness of conclusions reached might be implied but was not explicitly stated. However, given the subjective nature of much research there may not necessarily be correct answers as such. The focus, as regards assessing for critical thinking, appears placed more

objectively on the logic of the argument and not on whether the conclusions reached are right or wrong in the estimation of academics.

Critical thinking as argumentation involves being able to determine the validity of arguments as structured but also involves determining if the evidence supports the conclusions (Bailin et al., 1999a; Butler & Halpern, 2020; Davies, 2015). This too came out strongly in discussions with participants. If students are to win the argument on paper, they must substantiate their conclusions with suitable evidence. ‘Suitable’ means referencing apt academic sources at a minimum but more usefully critique of sources. Critique must progress beyond simply acknowledging or describing the views of others. Critique incorporates interpretation, analysis, evaluation and inference (linking with both the Delphi Project consensus statement (see Facione, 1990) and the cognitive skills of the critical thinker listed in the models of Davies (2015) and Halonen (1995). Arguments must be unpicked, findings evaluated, sources compared, contrasted and synthesised, relative strengths and weaknesses examined, information synthesised, relationships uncovered, tensions identified, assumptions questioned and positions challenged.

Exercising these skills will serve to evidence critical thinking for participants. Students as critical thinkers deconstruct the arguments of others as part of building their own arguments, structured and evidenced appropriately to convince (Buskist & Irons, 2008; Jones, 2004; Lundquist, 1999). These skills also operate in the construction of one’s own arguments, something that the critical thinking self-help guides focus much less on and arguably therefore give insufficient attention to (Moon, 2008). Critical thinkers develop their own viewpoints, informed from interrogation, analysis and evaluation of the ideas and arguments of others (Hammer, 2017; Jonassen & Kim, 2010; Luque, 2011; Ramage et al., 2009). That students can construct their own arguments was clearly important to participants as evidence of critical thinking.

7.3.2 Dispositions

7.3.2.1 Scepticism

Critically thinking students are unwedded to their prior beliefs, and should not accept information unquestioningly but exercise due caution in absorbing knowledge from others after carefully examining their reasoning, conclusions and evidence (Stanovich & West, 1997; Browne & Freeman, 2000). In practical terms, this can be evidenced through the undergraduate dissertation mechanism. However, for participants scepticism is not limited to challenging extant knowledge, cynically or otherwise. The student who is thinking critically moves beyond this, opening up and seeing the world in a more questioning way but also in a more personal way. This is perhaps more difficult to evidence through the undergraduate dissertation mechanism.

In addition, for participants sceptical students are constructively self-critical. They have the courage not only to question others but also themselves (Dominguez, 2019). This too links with the cognitive skills aspects of the Delphi Project consensus statement, which includes reference to “self-regulatory judgement” (Facione, 1990, p.6). Self-critics are self-regulatory, “brave enough to risk being wrong, and wise enough to realize that much can be learned from errors and failed solutions” (Nelson, 2005, p. xiv; cited in Davies, 2015, p.53). Critical thinkers in any context, not localised or limited to undergraduate students, should be alive to deficiencies in their own thinking through constantly testing their rationality against differing perspectives (Brookfield, 1987).

7.3.2.2 Open-mindedness

Participants are looking for students to adopt a position, having weighed up the alternatives, and argue effectively for that position. This requires open-mindedness, but also fair-mindedness and organised thinking. Critical thinkers are thoroughly disciplined in their approach to thinking through information, arriving at reasoned determinations (James et al., 2010). Students should not be impulsive or precipitate in taking positions therefore. A degree of prudence is required. This is about students working through multiple perspectives, recognising the situated complexity of interpretations that are not necessarily right or wrong, addressing the positions taken by commentators, and building their own positions. Being open to different interpretations, different perspectives, different approaches, enables the student as critical thinker to have

greater appreciation of context and a deeper understanding of the position ultimately adopted (Browne & Freeman, 2000).

7.3.2.3 Flexibility

For participants, flexible students are unfazed by problems with dissertations. Perhaps students have scheduled interviews but these have been cancelled because the intended interviewees are no longer willing. Perhaps large-scale surveys have yielded very few responses, too few for the results to carry the necessary statistical efficacy. Perhaps students could not feasibly collect data on a particular variable required for a mathematical model. These are examples of practical problems that a flexible and self-regulatory student will recognise, rationalise and suitably overcome. After all, dealing with the pitfalls of research that is "*messy*" (*Evelyn*) is a rich part of the research process for participants. They expect students to be self-critical of their own research approaches (sceptical), alert to problems and adaptable in finding solutions. Flexibility is needed therefore, as are complex skills of problem solving, linking with the cognitive skills elements of the Delphi Project consensus statement and the critical thinking in higher education models of Davies (2015) and Halonen (1995).

7.3.3 Originality

Originality manifests in creativity and in contributing to knowledge (see, for example, Anastasiadou & Dimitriadou, 2011; Halpern, 2014). Participants concur, although this is moderated for undergraduate students at their stage of academic development.

7.3.3.1 Originality - Creativity

Creativity, in the context of discussions in and around undergraduate dissertations, relates to students thinking differently about their research topics and the problems at hand; evidenced by some element of novelty, for example by students going beyond the norm in their approaches to research and thereby exceeding participants' expectations. For participants, agreeing with Dumitru (2019), such novelty can come from anywhere in a research dissertation, including its context, subject, research questions and methods. This is likely to involve some imagination as well as taking on some risk (Frick, Albertyn & Bitzer, 2014; Sternberg & Kaufman, 2010). Perhaps understandably, final year undergraduate students may want to play it safe when it comes to dissertation research, picking from a limited range of topics and devising formulaic questions with a view to pursuing an acceptable grade given operational constraints. Economists might label this

‘satisficing’. Perhaps it is unfair to complain that many students seemingly pick from the “*low-hanging fruit*” (*Olga*), their research questions lacking in creativity. In any event, creativity could equally be apparent in the creativity with which the questions are pursued. As said, even a study based on a well-trodden path with familiar questions can be interesting and exhibit creativity in the way in which the student has gone about fulfilling these.

7.3.3.2 Originality - Contribution

What is different about a dissertation study and its findings? What is unique? What insights are offered? What do we now know from a dissertation that we did not know previously? What is valuable about it? What is useful about it? What, for example, are the implications for policy and practice? Irrespective of whether a contribution is theoretical and/or practical, I see this as akin to building a wall, in this case of knowledge. Appreciably, undergraduate students are unlikely to generate new breakthrough knowledge, i.e. knowledge that advances the disciplinary field (Baptista, Frick, Holley, Remmik, & Tesch, 2015). However, the undergraduate student can contribute by adding a brick or two to our existing knowledge. Sometimes there is a gap in the wall that the student can help to fill, or it may simply be a case of adding at the top. Whichever, students are effectively then part of the academic conversation, building upon the work of others and making their own contributions.

In simple and practical terms, an undergraduate student can contribute through finding and identifying a gap in the literature. The student assesses the current state of knowledge, determining what is known, what is missing and what needs to be known, pointing out that the literature has certain limitations, identifying where the literature is deficient and where further research is needed to address this. In a sense, the student is finding the novel through determining what is and is not there.

Other ways of contributing include being innovative with approaches and methods (Lovitts, 2005), or finding the novelty in contexts, participants and the subject of the research (Dumitru, 2019), resulting in incremental additions to what we know. It is also worth noting that business research is often situated in real-world situations in which protagonists are looking for practical solutions to problems rather than theoretical propositions. Contributions can be made, therefore, by offering valuable and useful suggestions that are situationally appropriate or meaningful, e.g. for professional

applications or policy implications (see Bennich-Bjorkman (1997); Beghetto (2013); Halpern, 2014; Halpern & Sternberg, 2020; Pope, 2005).

7.3.3.3 Originality - original thinking

Participants are interested in what students think and have to say and ideally this should come through a dissertation in some way. In essence, if students own original thinking were to come through their work, evidencing application of their intellect to the problems under investigation, demonstrating their understandings, impressions, perspectives and positions, then that could be perceived by two-thirds of participants as offering some original perspectives and hence present as more critical.

7.4 RQ3: What are the challenges to achieving consistency when assessing for critical thinking in undergraduate students' dissertations?

7.4.1 Challenge 1: differing expectations of originality

7.4.1.1 Originality as contribution to knowledge

Opinion amongst participants was certainly divided on whether originality ought to be expected of final year undergraduate students. The division sits with differing expectations of educational levels. That originality is required at doctoral level is a given and collectively understood in higher education (Baptista et al., 2015). Participants would not dispute this. Originality is what gives a doctoral thesis the necessary 'doctorateness' (Wellington, 2013). It is certainly worth noting at this point that some of the literature I have cited in and around originality clearly has doctoral research in focus (see, for example, Baptista et al., 2015; Frick et al., 2014; Lovitts, 2005; Wellington, 2013). At master's degree level, academics who participated in the study by Bettany-Saltikov et al. (2009) felt that originality should be there and explicitly assessed for and that should affect the mark, even though a requirement for originality was not part of the University's applicable assessment criteria. It is worth noting that those participants numbered only four and although coming from different disciplines these did not include any business disciplines. Moreover, the focus of my research was on final year undergraduate students not masters or doctoral students. These are at a comparatively early academic stage.

Hilberg (2010, p.58) describes originality meaning contribution to knowledge as a “time-honoured scholarly ideal” but questions its applicability to undergraduate students. Yet a third of commenting participants, what I would regard as a significant minority, feel that originality should be there even in undergraduate dissertations and that this would necessarily evidence critical thinking. The feeling is that without this a dissertation will be short of real understanding and value as a piece of scholarly work. One participant would go so far as to suggest that the assessment criteria calls for originality for higher grades, although this is not corroborated. Others would agree that an original contribution would evidence critical thinking, but would not expect it of undergraduate students.

Hilberg (2010) would distinguish originality as it is commonly understood from independent work that offers a scholarly contribution. The latter requires of undergraduate students some independent working, but does not require innovation *per se* or that the student’s work should break new intellectual ground. I think Hilberg recognises that undergraduate students are highly unlikely to make such a breakthrough, given their academic stage and the generic approaches to research most often pursued. For Hilberg (2010) it is the independent nature of the research undertaken by undergraduate students that is important, not *originality*, arguing that such expectations are entirely questionable. There are always exceptional students of course, hence, there is recognition that originality if observed would certainly impress.

Of concern is that differences across participants here feed into assessment practices leading to inconsistencies. Assessors who expect originality will actively seek it. Should they not find it then that is likely to lower their impression of a piece of work and hence the grade awarded. The converse is likely true for assessors with no such expectation. The absence of originality will not affect a grade if unsought; however, its appearance would elevate the grade.

7.4.1.2 Originality as methodological creativity

If creativity is doing something unusual per Halpern & Sternberg (2020) then a practical way for undergraduate students to demonstrate this in dissertations is through their research methodologies and methods. Participants see worth in students reaching beyond what they have been taught in class to engage with unfamiliar methodologies and methods. Participants value the bravery of students in engaging with the added complexity. Always assuming such engagement is successful. Otherwise, if poorly executed this will tend to detract, doing more harm than good.

Participants have their individual predilections for certain methodologies and methods. Several participants had quite definite views on where they would place themselves on a quantitative versus qualitative research divide driven by background factors such as disciplinary and journal preferences. It is conceivable that qualitative approaches might afford students more opportunity for demonstrating their critical thinking, as Olivia suggested. I can see from my own experience of supervising undergraduate dissertations that qualitative approaches would tend less towards the formulaic than quantitative approaches, particularly with reference to replicative-style studies, which are predominantly quantitative in approach. Of course, colleagues who favour quantitative approaches may well disagree.

Regardless of methodological preferences, all participants are looking for competent execution of research. However, that Zhao individually believes methodological distinctiveness is required of undergraduate students, and that this would influence his assessment, is of concern. This is not supported by the assessment criteria, nor by the majority of participants who favour methodological fit over creativity. Zhao may simply have been verbalising what he believes to be commonly understood. However, work with what you have, use the right tools or techniques for the job, and do not try to be creative for the sake of it, appears to be the majority view. Instinctively I would locate this within a larger debate around supervisors/assessors' methodological preferences. However, Zhao volunteered his comments tangentially in the course of the conversation. These were not explored in detail with him, nor raised by other participants, so I am unable to conjecture that this points to a larger issue for consistency. On the evidence of the dataset, his appears a lone voice limiting the resultant inconsistency within the local assessment community, but inconsistency that is nonetheless apparent.

7.4.1.3 Description and replication (as unoriginal)

Participants deem ‘descriptive’ work unoriginal. Descriptive work is generally basic or simplistic, perhaps tending to the superficial, lacking in depth of thought, possibly unsophisticated in research terms, would report and describe more than analyse, interpret and synthesise, would fail to think through implications and develop contributions. I feel that such work could point to students operating more at the less developed stages of the epistemological development models of Perry (1970) and Baxter Magolda (1992) and hence less critical thinking.

Replication-style research is not automatically descriptive in this sense. It is, however, unlikely to be perceived as creative and hence, original. Bennich-Bjorkman (1997) would describe replication of research as lacking originality due to the apparent absence of novelty. Similarly, participants perceive replication-style studies to be inherently unoriginal because the student follows the blueprint of a prior study. In fairness, a small majority of participants were positive of replication-style research, which seemingly lowers the risk for students in submitting work of at least passable standard. In addition, replication-style research can arguably aid knowledge development, helping to support the reliability of research findings from and the trust placed in prior studies (Baptista et al., 2015). This type of research can provide a contribution therefore, offering originality and hence evidence of critical thinking.

Why students adopt this type of approach, apparently in large numbers in disciplines such as Finance, could conceivably come down to their motivation for undertaking research. If they view the dissertation as simply a task that they need to get through and finish within a limited timeframe, rather than something they care passionately about, then doing something that has been done before following a readymade template makes a lot of sense. This is especially so if, as suggested by Milo and Evelyn, students can be successful with this. There is nothing to prevent students from producing technically sophisticated work when adopting this route, which is highly valued by Oliver for example. If students can achieve a first class grade from an approach which arguably demands less initiative, creativity, originality, etc., because there is a blueprint to follow, and which carries less risk with the likely grade, then why wouldn’t they take this approach? To do otherwise would seem irrational in Economists’ terms. However, with the approach described as “*boring*” (*Olga*) and a “*dissertation by numbers*” (*Florence*), the positive views offered by Milo and Evelyn are unlikely to hold across the board.

Perhaps then of concern for assessment consistency the risks for students appear less to do with adopting a replication-style approach *per se* and more to do with variability in the cynicism of individual academics for this type of approach and any consequent impacts on their assessment outlooks.

7.4.1.4 Fresh literature at the findings and analysis stage (evidencing original thinking through making links with other areas of literature)

Whilst participants generally expressed the need for fit and alignment throughout a dissertation, in particular between the findings and the literature review, a small minority of participants were of the opinion that excellent students should be bringing in fresh literature at the findings stage, i.e. literature not covered in the earlier review, as evidence of original thinking.

It occurs to me that on the one hand a dissertation could be neatly wrapped with findings and analysis clearly referenced back to the literature review. On the other hand, a dissertation could bring in additional literature to demonstrate the development of the thinking and potentially some originality of thought through the analytical process. In reality, it is unlikely that research is so procedurally formulaic that the literature review will encompass everything that is necessary to help explain the findings. Some participants find the neatness plausible, depending on the nature of the study and the research approach taken, for example in certain quantitative traditions where theories are operationalised and prescriptively tested. Others, as has been said, find research to be quite "*messy*" (*Evelyn*). Take, for example, a more open-ended, explorative type of research study, perhaps focussed on gathering and analysing qualitative data, in which new insights can be gleaned from iterative interpretation of the data unconstrained by the chosen parameters of the preceding literature review. Here, participants like Arthur would argue for extending the literature at the findings stage if something has emerged. In so doing, students are showing that they have then thought about it and have related it to other things that they have read, or have sought new avenues of literature to help explain what they have found, demonstrating some original thinking and hence their critical thinking.

What matters most for this thesis is that this points to differences in approach when it comes to assessment. All participants expect congruence between the literature review and findings. For the majority of participants that would be that. Conversely, for a minority of participants an extension of the literature base in response to findings would evidence some original thinking on the part of the student, for which credit would be given. It is conceivable that, subject to preference, the addition or otherwise of fresh literature at the findings stage could delight or jar for participants, which could then affect the grade and challenge consistency.

7.4.2 Challenge 2: holistic versus additive approaches to assessment

Participants' descriptions suggest two categories of approach when assessing undergraduate dissertations: holistic and additive (akin to "impressionistic / holistic" and "analytic (weighting method)" per Pathirage et al., 2007, p.274). In both cases, a dissertation is seen as constructed from different parts, but the way in which a final grade is attributed to the piece differs.

Holistic assessment practices are rife in higher education (Bloxham et al., 2011). With a holistic approach, the assessor reads the dissertation as a whole and forms an overall impression against the assessment criteria, mentally balancing all aspects, allocated as a single grade. This offers maximum flexibility for the assessor (Pathirage et al, 2007) and benefits the assessment by suspending assessor judgement until the end when the piece can be considered in its entirety. Nightingale et al. (1997) further claimed that a process of holistic assessment that encompasses complex assessment criteria appropriately aligns with the assessment of higher-order graduate skills such as critical thinking. However, the difficulty lies in that mental balancing act. Studies in the field of cognition have shown that characteristics inherent to complex decision-making cause inconsistencies in academic judgement (Brooks, 2012). Assessors must engage in complex mental manipulation of a set of assessment criteria, criteria that can merge and interact (Sadler, 2009). The process is one of juggling multiple criteria concurrently in multi-layers reducing to a single representation (Bloxham et al., 2011). As human beings our capacity to process 'chunks' of information simultaneously is limited, and this practice inevitably places considerable demands on our cognitive load (Miller, 1956; Yorke, 2011). Hence, the cognitive challenge of assessing against and across several overlapping criteria is severe and assessors understandably find this difficult (Shay, 2004).

With an additive approach, an assessor applies criteria to distinct parts of a dissertation, e.g. literature review, methodology and methods, findings and analysis, etc., allocating a series of grades to individual sections and summing to an overall grade. I anticipate this will lessen assessor subjectivity because of the reduced cognitive burden compared with a holistic approach, and through building a more detailed picture of the grade awarded. Whether this operates so is open to debate, particularly given the added subjectivity of some participants apparently devising and applying their own additive assessment schemes. Clearly, this would work against the added standardisation which Pathirage et al. (2007) claims an additive (or analytic) approach provides. Allocating grades analytically against each criterion through separate qualitative judgements is also impractical because it is time consuming (Hornby, 2003; Sadler, 2009; Yorke, 2011). Unsurprisingly therefore this practice was not observed in a study by Bloxham et al. (2011).

In reality, the choice of approach is not a binary one of holistic versus additive. Participants, based on their descriptions, could be placed in either category, or in some cases they would blur the distinction by engaging in a hybrid approach of adding towards a summed grade but also forming a holistic check, with one acting as confirmation of the other. This presents its own difficulties. What happens if, as suggested by Sadler (2009), a discrepancy exists, i.e. where the sum of the parts is more or less than the assessor's holistic assessment of the worth of a piece of work? Research suggests that an assessor faced with such a discrepancy may trust in the latter and override the former (see, for example, Baume et al., 2004; Grainger et al., 2008). Perhaps they may apply a "fudge factor" to overcome perceived deficiencies in the criteria (Walvoord & Anderson, 2010, p.119). This can only add to the subjectivity, as does each participant apparently approaching the task of assessment in a way that suits them individually, meaning that approaches across participants are not consistent.

7.4.3 Challenge 3: sharing understandings and calibrating assessment

In principal, mandated application of official assessment criteria, and internal moderation of assessment via double marking processes, serve to facilitate the sharing of understandings and the calibration of grades. However, there are questions as to their effectiveness, both in the literature and in relation to the practices discussed and observed at the host business school.

7.4.3.1 Assessment criteria

Half of participants think that assessing for critical thinking in undergraduate students' dissertations relies on professional judgement, combining knowledge and experience with gut instinct. Professional judgement is an important resource; however placing sole reliance on this invites inconsistency across assessors. Criterion-referencing practice serves to mitigate this, giving assessors an applicative base from which to judge students' work. 'Standards-based assessment rubrics' (grade descriptors) provide qualitative descriptions of what is expected at different levels (Kite & Phongsavan, 2017). The criteria set general expectations applicable across the board, seeking to move assessors away from personal biases. I am told that being 'critical' is integral to the criteria at the host business school, particularly as the grade scale progresses. However, critical thinking is an outcome with high-level cognitive demands that is not easily defined or measured. Whilst it is plausible to construct criteria that account for critical thinking, it is hugely difficult to interpret these on a consistent basis (Neil et al., 1999; Woolf, 2004). What is 'critical' is not precisely specified in the criteria and I bear witness to what participants' view as constituting critical thinking in students, and what evidences this in students' work, can and does vary. This mirrors the literature, which would define critical thinking variously with no clear academic consensus on what it is and how it should be specified (see, for example, Baril et al., 1998; Bissell & Lemons, 2006; Butler & Halpern, 2020; Harrell, 2011; Kek & Huijser, 2011; Lok et al., 2016; Moore, 2013; Nicholas & Raider-Roth, 2011).

Consistent application of assessment criteria is important in propagating shared understandings between assessors to reduce bias and achieve transparency and consistency by and across assessors in the quality assured accountability regime of UK higher education (Bloxham et al., 2016b; Menendez-Varela & Gregori-Giralt, 2016). Most participants agreed with the use of assessment criteria, as mandated by the host business school, seeing the value in doing so for assessment consistency both by individual assessors over time and across assessors.

Predictably, the picture is unclear when it comes to the effect of the use of assessment criteria on assessment reliability. Criteria are said to provide a supportive crutch to lean on, particularly for inexperienced assessors (Jansson et al., 2019). Yet assessors' understandings, interpretations and the ways in which they combine criteria can and do vary (Sadler 1987, 2005, 2009). Evidently, assessors may still arrive at different grades for the same piece of work whilst applying the same criteria (Grainger et al., 2008). Some studies have shown improvements in assessment reliability following the introduction of criteria, if only small in extent (see, for example, Jonsson & Svingby, 2007; Peeters et al., 2014). Others would contend otherwise (see, for example, Baume et al., 2004; Price, 2005; Yorke, 2008). Studies on both sides of this divide are premised on consistent use of criteria. Here, however, differences were apparent in how participants actually use the criteria. This ranges from systematically checking against the criteria, to having it to hand as a reminder reviewing it at intervals, to applying it in uncertain cases only, for example borderline grades, to applying it after the event as a final check on an already determined grade. It is unclear what difference this might make to grades and consistency of assessment, although I submit it is unlikely to aid the latter. Ultimately, the criteria are being used and whilst exact approaches may vary, each as a minimum employs the criteria as a useful check on professional judgement. However, I would question the effect on assessment consistency given the evident inconsistencies in how the criteria are actually applied.

Going untested here also are questions of interpretation of criteria. Criteria are claimed to be of limited power as agents of consistency; their meaning is 'fuzzy' and so not effectively or reliably communicated (Bloxham, 2009; Bloxham et al., 2016a; Ecclestone, 2001; Price & Rust, 1999; Sadler, 2009). Some contend that the way to deal with this is to work on the detail, to render criteria as clear and explicit as possible in order to improve collective understanding (Bloxham et al., 2016). Others question our reliance on criteria, claiming that this is unlikely to lead to consistent decision-making because

criteria can never be sufficiently clear or explicit enough to eradicate inconsistencies in interpretation (Bloxham et al. 2016; Sadler, 1987). Exercise of professional judgment is inescapable, if only to resolve this 'fuzziness' (Bloxham, 2009; Yorke, 2011). Criteria are interpreted through the lens of personal experience. Interpretations vary and are rarely incontestable (Bloxham et al., 2016). Hence, even consistent use of criteria, which I have not witnessed from participants, would not necessarily achieve consistency of assessment because of the likelihood of differing interpretations.

Use of criteria has then its problems. Non-use of criteria, or substitution by one's own criteria, are likely to be of even greater concern. Three participants admitted non-use, two of whom claimed confidence in their knowledge and understanding of the criteria held from long-standing use. They did not have an issue with using the criteria *per se*, rather they did not feel the need to refer to it whilst assessing undergraduate dissertations because they considered that they had sufficiently internalised it. However, I might question the accuracy of their respective recollections of the criteria. I might also question if these are up-to-date with any changes made over time.

One participant, Olga, seemed 'off on a frolic' to use a legal term in apparently basing her assessments on a comparative basis across scripts with standards wrought from her experience. This suggests more of a norm-referenced approach, which I would think sits ill with the modern paradigm of quality-assured accountability, by which a student's work should be assessed strictly on its own merit with reference to assessment criteria and without reference to the student's peers or his or her past work (Sadler, 2009). That said there is a view that norm referencing does in any event tacitly inform assessment practices (Yorke, 2011). Assessors will inevitably compare, drawing on their knowledge of a student's work and across students in making judgements. Bloxham et al. (2011), for example, observed assessors explicitly comparing students' work whilst engaged in an ostensibly criterion-referenced assessment exercise, although the nature of the study – participants were required to grade two submissions of the same assignment – may have encouraged norm referencing through comparison. Often assessment criteria is itself constructed implicitly from norms and normative knowledge, and, given that criterion referenced assessment strategies are moored to a specific context, criteria are normatively interpreted in that context (Lok et al., 2016; Orr, 2007; Sadler, 2005). The reality is that most assessors, consciously or no, are subject to a degree of normative influence promoted through normative disciplinary practices (O'Hagan & Wigglesworth, 2015; Yorke, 2011). In effect, assessors employ a mix of criterion referencing and norm

referencing in allocating grades in the expectation that a grade can be determined with reference to the former whilst respectful of the convention of the latter (Lok et al., 2016). This is particularly observable with 'difficult' decisions such as borderline grades (Grainger et al., 2008). From this, it is possible that Olga is not alone in apparently employing a degree of norm referencing, although as far as the dataset goes only Olga gave any indication of this.

It is also unclear whether or to what extent the official criteria are included in Olga's assessment practice. "*Not very strictly*" was her response when questioned as to her use of the criteria. It is possible that the criteria have informed Olga's standards if she has used them in the past. I anticipate also that working with colleagues in the local assessment community, many of whom will be using the criteria, will have informed Olga's standards. These, although apparently individualised, may or may not in effect deviate substantially from others because of such influences. However, such tacit or internalised criteria are a particular barrier to consistency across assessors because they are owned by an individual, formed from their own preferences, not necessarily drawn from official criteria or shared with colleagues (Baume et al., 2004; Grainger et al., 2008; Hunter & Docherty, 2011; Price, 2005; Read et al., 2005; Shay, 2005; Webster et al., 2000).

Subjectivity is claimed to be inherent to assessment and therefore inescapable (Clegg and Bryan, 2006). As if to prove this point, five participants made reference to adopting criteria of their own devising, mentally substituting this for the official criteria. Again, as with Olga, whilst conceivably these may be shaped by the official criteria in tandem with interactions with colleagues this was not explored. That some participants are apparently applying their own personal standards or tacit or internalised criteria would inevitably point to variability (see, for example Baume et al., 2004; Bloxham, 2009; Bloxham et al., 2016; Ecclestone, 2001; Hunter & Docherty, 2011; Price, 2005; Price & Rust, 1999; Orrell, 2008; Read et al., 2005; Shay, 2005; Smith & Coombe, 2006; Webster et al., 2000; Woolf, 2004).

7.4.3.2 Internal moderation (double marking)

Shay (2004, p.307) calls assessment in higher education a “socially situated interpretive act”. Within a socio-cultural framework, learning is jointly constructed not passively acquired. Shared understandings develop from members operating closely together in a local context, sharing reflections and engaging in structured conversations (Bloxham, 2009; Bloxham & Price, 2015; Maxwell, 2010; Menendez-Varela & Gregori-Giralt, 2016; Stupans, March & Owen, 2013). The meanings attributable to assessment criteria, for example, can only be operationalised in a specific context, with a specific group of assessors who calibrate collective understandings through a process of socially situated co-construction and interpretation informed by the tacit knowledge of members (Baird, Greatorex & Bell, 2004; Bloxham & Price, 2015; Bloxham et al, 2015; O’Donovan et al., 2004; Rust, O’Donovan and Price, 2005; Shay, 2008).

The host business school has a system of internal moderation via double marking, whereby two assessors independently assess a dissertation and then meet to meet to compare, discuss and agree upon a final grade. This, as with any system of internal moderation, plays a fundamental role in quality assuring assessment practices (Beutel et al., 2017). Ecclestone (2001) refers to the dangers of unconscious bias that each assessor may hold built over time in post. The aim of internal moderation is to improve consistency (Grainger et al., 2016; Sadler, 2013). Discussions between assessors over standards, if collegial, could serve to minimise such biases (Yorke, 2011). Moderation meetings between assessors bring out comparative views, helping assessors to get a sense of what they are looking for with respect to the assessment criteria, to understand what they are picking out as critical or uncritical in dissertations, to get a feel for where they and their expectations are relative to others. These discussions, held within local disciplinary assessment communities, are key to developing shared understandings and so aid assessment consistency (O’Hagan & Wigglesworth, 2015; Price, Carroll, O’Donovan & Rust, 2011). This then is a key mechanism for sharing knowledge and developing shared understandings amongst assessors (Grainger et al., 2016). Participants confirmed this by two-thirds majority. Participants also confirmed that questions over critical thinking in undergraduate dissertations are invariably worked through in meetings between assessors.

Moderation thus forms an important element of socialisation through which assessors collectively cultivate local standards (Reimann et al., 2010). Meanings are constituted through shared practice and dialogue within the particular social context (Bloxham et al., 2015; Shay, 2004). Members of the local assessment community learn from each other. In particular, newer members learn from more experienced members, evolving their understandings, calibrating their expectations and developing a 'feel for the game' (to paraphrase Bourdieu). Calibration of standards is not a technical matter but a shared interpretive act bounded by and limited to a local context (Bloxham et al., 2015). It follows that the longer a member is a part of this community, interacting with colleagues in various ways but particularly through the moderation mechanism, the more understandings of critical thinking will converge and the more grades will calibrate. Through moderation, assessors gradually absorb criteria that are implicit to their community and necessarily rely upon the interpretive support provided by their community to form understandings (Jawitz, 2009; O'Connell et al., 2016). This necessarily assumes a degree of stability in that local assessment community at the host business school, which I have observed from participants backgrounds especially at the experienced end.

In theory then, internal moderation via double marking should generate shared understandings and calibrate dissertation assessment. That said there are I think impediments at the host business school to how this works in practice. Let us take, for example, a limitation on sharing. For the most part, an assessor is paired with only a limited number of other assessors, subject to volume. The dissertation module leader decides the pairing allocations, not assessors, although they may make representations to the module leader as to with whom they would wish to be paired. Occasional representations aside, I observed the allocation process to be incremental, i.e. it was set some years ago and has largely remained static from year to year, incrementally adjusted by the module leader as members join and leave. This limits the degree of mixing between assessors and hence the degree of sharing of understandings and calibration of assessment outwith small pockets of assessors. Understandings and calibration are not necessarily translating across disciplinary and school communities of assessors therefore. The effectiveness then of internal moderation as a mechanism for reducing variation in assessment may be questioned.

7.5 Brief summary

In this discussion chapter, I returned to the three research questions I posited in the introductory chapter, namely:

RQ1: How do business school academics conceptualise critical thinking?

RQ2: How are business school academics' conceptualisations of critical thinking operationalised through assessment practices?

RQ3: What are the challenges to achieving consistency when assessing for critical thinking in undergraduate students' dissertations?

I began, in answer to RQ1, by presenting the construct of critical thinking for participants as business school academics as a composite of skills (of argumentation), dispositions and originality, drawing on my findings from Chapter 5. I followed this, in answer to RQ2, with consideration of how this concept of critical thinking is operationalised by participants when they assess undergraduate dissertations, examining what participants look for from students as evidence of their critical thinking, again drawing heavily on my findings from Chapter 5 but also aspects from Chapter 6. Responding then to RQ3, I discussed the apparent challenges to assessment consistency born of inconstant approaches and less than effective institutional mitigating mechanisms, drawing on my findings from Chapter 6.

I now look to conclude this thesis in the following chapter. I begin with an overview from initial premises through to a summary of key findings related to the research questions, RQ1-RQ3 inclusive, before developing some points of theoretical and practical significance arising from this research. I then make recommendations for extension through future research in advance of some brief closing remarks.

8. Conclusions

8.1 Introduction

In this, the eighth and final chapter, I draw this thesis to a close. I begin with a brief overview of this thesis, from the research rationale drawing on aspects of the critical thinking and assessment literature building to three key research questions, RQ1-RQ3 inclusive, then to how the research proceeded through to the development of themes in answer to those research questions. Key findings, detailed in Chapters 5 and 6 and further discussed in Chapter 7, are then summarised, following which I draw the significance of this thesis in terms of theoretical contributions and practical implications for the host business school and potentially beyond. Finally, I offer suggestions for further research to extend the work of this thesis before concluding with some brief closing remarks.

8.2 Thesis overview

I introduced this thesis in Chapter 1, outlining its inspiration, parameters and setting. Critical thinking may be a defining concept of higher education, and a term used pervasively, particularly with respect to assessment, yet it is a difficult term to get to grips with lacking as it does clear definition, specification and conceptualisation. High stakes assessment is a central function of higher education and significant efforts have been made to improve fairness in assessment through greater accountability and transparency, yet research continues to demonstrate difficulties and variability, particularly with regard to assessment of complex aspects such as critical thinking.

It seems that critical thinking is of huge importance to assessment in higher education. However, what exactly is critical thinking in this context? How is it discerned? How is it assessed in undergraduate students' work? What are the challenges to achieving consistency in this? I sought to investigate this through a case study of a UK university business school, building a participant sample of academics across five business-related subject disciplines who would have an understanding of critical thinking in context, who would have experience of assessing for critical thinking in undergraduate students' dissertations and who would thus be positioned to offer relevant insights.

Data collection produced transcribed data from twenty-one semi-structured interviews, the format of which included an element of document elicitation involving assessed undergraduate dissertations. From detailed thematic analysis, following the approach of Braun and Clarke (2006), I developed two distinct themes: (1) Facets of an undergraduate student as critical thinker (detailed in Chapter 5); and (2) Approaches to assessment: convergence, differences and mitigation (detailed in Chapter 6). Together these themes capture the essence and key aspects of the dataset as I have interpreted them. I then examined these themes set against RQ1-RQ3 inclusive in relation to a literature base of critical thinking and assessment in higher education (detailed in Chapter 7).

8.3 Key findings summarised

8.3.1 RQ1: How do business school academics conceptualise critical thinking?

Theme 1 portrays how participants in this study conceptualise critical thinking in their particular higher education context. This concept is comprised of skills (of argumentation); originality (creativity, contribution and original thinking); and headline dispositions amalgamated from smaller groupings (scepticism, open-mindedness and flexibility).

Firstly, skills of argumentation are central to how participants' conceptualise critical thinking. This mirrors the first part of the Delphi Project consensus statement (See Facione, 1990) and follows the argumentation as critical thinking approach of the study skills texts (see, for example, Cottrell, 2017) which is prevalent in higher education. These critical thinking skills manifest for participants in the way that undergraduate students structure their arguments and support these with evidence, i.e. they look for arguments that are structurally robust and evidentially sound as indicative of the exercise of critical thinking. Students can demonstrate their critical thinking through constructing valid, believable, logically structured and reasoned arguments that avoid contradictions, and which build from legitimate premises through to sensible and substantiated conclusions, supported by suitable evidence such as referencing, comparing and contrasting, critique and synthesis of sources.

Secondly, in tandem with the necessary skills, students who are critical thinkers will exhibit certain dispositions. Theme 1 portrays a distilled localised list comprised of scepticism (incorporating curiosity and inquisitiveness); open-mindedness (incorporating prudence, tolerance of ambiguity, appreciation of individual differences); and flexibility (incorporating persistence and self-correction). Scepticism – students as critical thinkers take nothing for granted and question all. Open-mindedness – students as critical thinkers are curious and inquisitive, open to alternatives and willing to examine them, weighing up the debates before adopting justifiable positions. Flexibility – students as critical thinkers are not wedded to their beliefs, are willing and able to change views where appropriate, are unfazed by problems and are able to adapt.

Thirdly, for some participants, students who are critical thinkers will demonstrate some originality, be it creativity through innovation, novelty, etc., or contribution to knowledge whether theoretical, methodological, practical, professional or policy related, or original thinking evidencing application of their intellect, their understandings, impressions, perspectives and positions. However, that originality should be expected of undergraduate dissertation students was not universally accepted by participants. Opinion was very much divided on this. Originality would certainly impress and, whether expected or not for the level, originality was routinely discussed by participants and it does form part of the conceptualisation here of critical thinking.

Fourthly, I offer an addition to what has gone before in terms of conceptualising critical thinking in higher education more broadly, i.e. knowledge, comprised of subject knowledge and epistemological development. This is fundamental to the abilities of students to engage in critical thinking. Deep subject knowledge and the skills and dispositions indicative of higher stages of the epistemological development models of Perry (1970) and Baxter Magolda (1992). These models speak to what is and is not ‘critical’ in students’ work, reflecting participants’ expectations of final year undergraduate students. This is of particular relevance to assessment of critical thinking in undergraduate students’ work.

8.3.2 RQ2: How are business school academics' conceptualisations of critical thinking operationalised through assessment practices?

Critical thinking as concept is operationalised in assessment practices through what participants look for from students as evidence of their critical thinking. The findings in Chapters 5 and 6, further discussed in Chapter 7, consider this in the context of final year undergraduate dissertations, offering practical examples of what for participants would signal critical thinking and how and where such signals might be located.

Firstly, participants expect that the necessary skills of argumentation are put to use in forming cogent arguments that are well structured, with due attention to logic and reasoning, and backed by suitable evidence. Participants expect that students, as critical thinkers, are able not only to deconstruct the arguments of others, but also to build their own arguments, structured and evidenced appropriately to convince. Participants expect that students as critical thinkers will develop and promote their own viewpoints, informed from their interrogation, analysis and evaluation of the ideas and arguments of others.

Secondly, participants look for the necessary dispositional elements to be displayed, i.e. that students are sceptical, both of the world and of themselves; that students are open-minded, open to multiple perspectives and alternative arguments carefully weighed to arrive at their own defensible positions; that students are flexible, able to solve problems, overcome challenges and adapt their thinking.

Thirdly, that originality 'may' be expected of students, manifest in the exercise of creativity or in contributing to knowledge or in offering some original thoughts ('may' as this expectation is contested with respect to undergraduate students). Creativity relates to students thinking differently about their research topics and the problems at hand. This may be evidenced by some element of novelty, for example by students going beyond the norm in their approaches to research and thereby exceeding participants' expectations. Students can contribute to knowledge by identifying gaps in the literature, or by offering valuable and useful suggestions that are situationally appropriate or meaningful, for example professional applications or policy implications. Participants are interested in what students think and have to say and this can evidence some original thinking and hence critical thinking.

8.3.3 RQ3: What are the challenges to achieving consistency when assessing for critical thinking in undergraduate students' dissertations?

From Theme 2, the challenges to assessment consistency derive from apparent inconsistencies in participants' assessment practices, born from differences in how participants conceptualise and operationalise critical thinking (with reference to RQ1 and RQ2 above), and from personal choice of approach. Notwithstanding the efforts of the host business school to standardise and communicate the requirements for assessing dissertations, inexact conceptualisations of critical thinking and inconstant approaches to assessing dissertations were apparent. Differences here would point to the likelihood of assessment variation.

Participants would generally concur on the necessity for students to have the right skills (of argumentation) and possess the right dispositions (scepticism; open-mindedness; and flexibility) for critical thinking to occur. Participants would also generally concur that originality is a further important component of students displaying their critical thinking, but would disagree on their expectations vis-à-vis undergraduate students. For practical purposes, with final year undergraduate students in mind, differences were evident in whether and/or to what extent originality is required; whether and/or to what extent methodological creativity is required; whether and/or to what extent replicative-style research is acceptable; and if fresh literature should be brought in at the findings and analysis stage of a dissertation evidencing original thinking. Differences of opinion here manifest in differing expectations of undergraduate students again pointing to the likelihood of assessment variation.

The host business school expects assessors to take a holistic approach to the assessment of undergraduate dissertations. This is apparent from communication mechanisms, i.e. instructions given to assessors through the dissertation module handbook and an annual communications meeting; also through the stipulated application of official assessment criteria to a dissertation as a whole and not in parts. Nevertheless, individual approaches varied. Some participants employed a holistic approach, others additive and others still a combination blurring the distinction, borne of individual preference. Most but not all participants would agree with the use of the official assessment criteria. However, where some participants would claim to apply the criteria religiously, others would claim not at all (substituting their own criteria, or in one case comparing across scripts), some in part, and some only as a post-hoc check,

corroboration or justification. Even those participants who advocate applying the criteria in full may not necessarily interpret the wording of the criteria in quite the same way, pointing to potential variation. The evident lack of consistency of approach, coupled with differing interpretations of criteria and ways of applying the criteria, would seem to invite variation.

The double marking internal moderation mechanism of the host business school potentially has a major role to play in facilitating shared understandings, in filtering approaches, and in calibrating assessors' expectations within the local assessment community, in order to offset inconsistencies in interpretation and approach and thereby reduce the likelihood of assessment variation. This is subject to the ongoing stability of the academic faculty, and given enough time for calibration to occur. However, the internal moderation process as enacted at the host business school is subject to deficiencies that are likely to reduce its effectiveness. For example, in the relatively static pairings of assessors from year to year and concomitant limitations on mixing and sharing within the local assessment community. The process does not necessarily fix the causes of variation therefore.

8.4 Significance of the findings

8.4.1 Theoretical contributions

Clarification of how participants as business school academics conceptualise critical thinking as a set of skills and dispositions (RQ1), and understanding how they operationalise this through their assessment practices (RQ2), add to the extant knowledge on critical thinking in higher education.

Whilst skills and dispositions are confirmatory, originality is specifically foregrounded here unlike the listings in the Delphi Project consensus statement and subsequently developed models of critical thinking in higher education (see, for example, the models of Halonen (1995) and Davies (2015)). Participants see originality, and hence critical thinking, in students being creative, in students contributing to knowledge and in students offering some original thoughts. Whilst the presence of originality was not universally required of undergraduate dissertation students by participants, it was clear that originality does form an important part of the concept of critical thinking in the higher education business school context in which this research was situated.

A further addition to this contextualised conceptualisation of critical thinking is the suggested link between models of epistemological development and assessors' expectations of critical thinking in final year undergraduate students. This is premised on the levels or 'ways of knowing' in the models of Perry (1970) and Baxter Magolda (1992) speaking to what is and is not 'critical' in undergraduate students' work. I submit that these models, in capturing the epistemological development of university students, are reflective of academics' expectations of students and their work at different academic stages. Higher levels of these models reflect what participants in this study are looking for when assessing for critical thinking in undergraduate dissertations.

These aspects add to our theoretical understanding of critical thinking in higher education, albeit localised to a specific business school setting. However, potentially the findings may be transferable to other higher education settings including other business schools based on Stake's naturalistic generalisation (refer to s.4.6.5 and s.4.8.3 for further discussion of this point).

8.4.2 Practical implications

The practical implications of this thesis are twofold.

Firstly, I have sought to demonstrate how critical thinking as concept is operationalised by participants through assessment. I have shown what participants expect of final year undergraduate students as regards critical thinking and what they look for specifically as evidence of this when assessing for critical thinking in undergraduate students' dissertations. Business school academics can compare their own understandings with this, which could be confirmatory and/or illuminating. Business school students too can learn from this illustration of what academics are looking for as evidence of critical thinking in undergraduate dissertations, helping them also to develop the necessary 'feel for the game' (again paraphrasing Bourdieu).

Secondly, I have highlighted several challenges to consistency when assessing for critical thinking in undergraduate dissertations. These derive partly from individual variations in how critical thinking is conceptualised, for example differences as to what constitutes originality, and whether and/or to what extent originality is required; and partly from varying approaches to assessment, for example the choice of holistic, additive or a combination of approaches. Institutional mechanisms potentially have a major role to play in facilitating shared understandings of critical thinking as concept, and the operationalisation of this through assessment, serving to reduce variation. However, the mechanisms are fallible, which limits their effectiveness.

Critical thinking is at the heart of a university education, at least in the traditions of the Western world. We expect students to move beyond passive reception of that which is given as known through to active engagement in critical thinking. High stakes assessment is a central function of higher education and a crucial factor in determining future study and career opportunities (Lok et al., 2016; O'Hagan & Wigglesworth, 2015). Students who demonstrate greater critical thinking necessarily attain higher grades. It is crucial therefore that we get this right. I hope that the host business school can learn from these points to improve its assessment practices. More broadly, I hope, subject to the limits of transferability (refer s.4.6.4), that other higher education institutions, including but not necessarily limited to business schools, might find their own resonance with these points leading to action and improvements in their practices.

8.5 Further research

This research is localised to one business school within a UK university. As a single case study, this could be useful at the outset of theory generation (Benbasat, Goldstein & Mead, 1987) or, as here, for adding to our understanding of existing theory through clarification and augmentation. Appreciably the findings cannot be generalised in the traditional sense beyond the context of the case, although to repeat I hope that readers might find for a degree of transferability with respect to their own contexts through the description of the case and discussion of its findings. To aid this further, this thesis might be expanded in a number of ways. I offer the following by way of practical suggestions:

- This research could be augmented by additional primary research in the present case setting to enable more informed triangulation of findings; for example, detailed participant observation of assessment processes including double markers' meetings.

- The pre-interview questionnaire and interview schedule included questions on how participants had come to know what critical thinking is and how their understandings had developed over time. Participants did offer some interesting if occasional views on this. Overall, answers were sparse, insufficient for detailed findings and discussion I would say. Personally, I am interested in these aspects and feel that it would be worthwhile revisiting these questions with participants in more detail in a tangential study.

- This research could be followed by an action research project to build improvements into the assessment systems of the host business school and monitor success. Findings from such a study could have wider applicability and implications for assessment systems more broadly in higher education.

- This research could be replicated at the host business school with larger numbers of participants from each of the five academic disciplines of accounting, Accounting, Business and Management, Marketing, Finance and Economics. This may allow for more nuanced comparisons and findings on disciplinary lines, examining also the question of whether critical thinking is a generic skill or disciplinary-specific.

- This research could be replicated with student participants to generate comparative views of what critical thinking is and how it may be evidenced for assessment purposes. The absence of student voices is not a limitation of this thesis because the focus is firmly directed at investigating academics' perceptions. However, students' perceptions are also important, particularly given the suggested addition of knowledge incorporating epistemological development to the concept of critical thinking as assessed in higher education. Students' perceptions are similarly worthy of investigation and of comparison with the perceptions of academics.

- This single case study could be followed by a multiple case study, involving several business schools from a variety of universities in the UK and beyond. This research in part challenges and amends existing models of critical thinking in higher education. The findings could be examined in more detail through multiple cases.

- The existing interview dataset might be revisited with a new research question focussed on adding to the generalist vs. specifist debate, i.e. whether critical thinking is a generic skill or a disciplinary specific one, assessing for patterns along business school subject disciplinary lines in conceptualisations of critical thinking.

8.6 Closing remarks

At the outset of this thesis, I reported the personal inspiration behind this research coming from my own transition to the higher education sector from professional accountancy practice. I wanted to know if my understanding of critical thinking, the importance and substance of which I had come to know subsequently over several years as an academic, was individual to me or shared with others. Moreover, I wanted to know if understandings were shared across a business school, if differences exist and how this might influence assessment. I feel I have a better understanding from having progressed this research but also now greater appreciation of the complexity. There is much commonality but also differences such that whilst I have portrayed a conceptualisation that is held generally I recognise that there is no unitary definition or conceptualisation held across the host business school. Evident challenges to assessment consistency arise from this and other issues as discussed.

Overall, this research contributes to the discussion of critical thinking in higher education as concept, with specific reference to the host business school forming the case study. It further contributes by examining the concept of critical thinking as operationalised through the assessment of final year undergraduate dissertations, highlighting also the challenges of maintaining consistency notwithstanding mitigating institutional mechanisms.

I have offered some small theoretical contributions through a contextualised conceptualisation of critical thinking in higher education. This confirms what we already think we know in many respects but adds by highlighting the contested issue of originality for participants, and linking models of epistemological development now with assessment of critical thinking and not just the pedagogy of teaching critical thinking, which to my knowledge is hitherto untapped.

I have also suggested some practical implications. Firstly, in specifying what participants look for as evidence of students' critical thinking when they are assessing undergraduate dissertations. This could be of use to both academics and students. Secondly, in illuminating several challenges to assessment consistency, including differing priorities with respect to aspects of critical thinking, differing approaches to assessment, and the pros and cons of institutional mechanisms put in place to mitigate variation. These could be of use to the host business school and other higher education institutions in improving practices.

Limitations (refer s.4.6) reveal that this study, as with any piece of research, is not perfect. However, I have sought to mitigate for the limitations within the operational constraints to which I am subject, and I have sought to provide sufficient information to enable readers to consider the relevance of the findings for their own settings, potentially then informing decision-making, policy and practice. I further set out my own reflexive position on this research to allay concerns of subjectivity, which I have in any event embraced (refer s.1.3).

I hope that the findings of this thesis can be used to help improve the sharing of our understandings of critical thinking in higher education and for the betterment of assessment practices benefiting academics and students at the host business school and more broadly in higher education.

Appendices

Appendix A: Grade Descriptors for Undergraduate Programmes

[redacted]**Grade Descriptors for Undergraduate Programmes**

Class	Mark(%)	Descriptor
First	86-100	**Exemplary.** Exceptional work showing insight into the topic; reflects a complete grasp of knowledge and understanding. Such work is only rarely encountered.
	76-85	**Outstanding**. Comprehensive knowledge of the topic, showing depth of understanding with evidence of judgement in selection and critical analysis of relevant material. Logically structured and clearly written.
	70-75	**Excellent**. Detailed knowledge of the topic, with evidence of judgement in selection and critical analysis of relevant material. Well written with good structure. Minor errors acceptable if compensated by excellence in other areas.
Upper Second	65-69	**Very Good**. Displays good knowledge and thorough understanding of the topic with evidence of broader understanding informed by wider reading. Less critical grasp of the subject than evident in a First Class answer.
	60-64	**Good**. Reasonably good knowledge and understanding, but little evidence of critical assessment or analysis. Coherent presentation but less well structured than seen at higher grades.
Lower Second	55-59	**Adequate**. Sound general knowledge of the subject as taught but lacks evidence of broader understanding. Presented in a satisfactory framework with relevance to the topic retained throughout.
	50-54	**Fair**. Adequate, except that the work may be rather thin or unimaginative, missing some key points or lacking in clarity.
Third	45-49	**Weak.** Exhibits defects such as: factually correct, but at an elementary level or a narrow selection of material with significant omissions or significant errors of fact or understanding or muddled; lacking cohesion and direction, or a misguided selection of material.
	40-44	**Poor**. Typically includes several and sometimes significant defects and is thus barely acceptable. May include very short answers that nevertheless include key points.
Fail	35-39	**Very poor**. A very thin piece of work containing evidence of only rudimentary knowledge of the topic.
	30-34	**Extremely poor.** The work demonstrates little relevant knowledge and/or understanding of the subject.
	20-29	**Clear fail.** Work that misses major elements of the knowledge base. Deserves recognition for making an effort to answer the question or address the essay title, but shows very little evidence of knowledge or understanding.
	10-19	**Serious fail**. Significant inability to engage with the question or essay title. Marks are awarded within this range for overall presentation, the odd relevant word in context but negligible evidence of knowledge or understanding.
	0-9	**Outright fail**. Work of very little or no value, or disqualified due to lateness, plagiarism or other disciplinary offences.

- Where relevant, the effectiveness of oral presentation skills (for example focus, presentation structure, audience communication, visual aids) will be considered
- For quantitative work, the application of analytical techniques, in addition to attention to detail and interpretation of results where relevant, will be considered.

Appendix B: Participant Information Sheet

Research participant information sheet

Study title: A study of critical thinking: conceptualisation and operationalisation in assessment practices.

I would like to invite you to take part in a doctoral research study which is sponsored by Durham University Business School. Before you decide I would like you to understand why this research is being carried out and what it would involve. This information sheet explains the purpose of the study and what will happen if you participate. It also provides details about the practical arrangements for the study and how we will ensure that the information you provide is treated confidentially. Full contact details are also provided at the end of this document in case you would like clarification on any point or if you have any questions.

What is the purpose of this study?

Higher education students are ultimately differentiated by their performance in summative assessment, the results of which have ramifications for seeking employment or access to higher level education programmes. Evidence of critical thinking is generally required for higher-level attainment. This study aims to investigate how higher education tutors conceptualise critical thinking, to what extent the concept is shared and how tutors apply the concept when assessing dissertations.

Why have I been invited?

The study is focused on higher education tutors who supervise and assess dissertations.

Do I have to take part?

No, your participation is entirely voluntary. If you decide to participate, you are free to withdraw at any time without any negative consequences for you.

What will I have to do?

Simply confirm your agreement to take part in the study directly with the researcher, who will then get in touch to make the interview arrangements. The researcher's full contact details are also provided at the end of this document. If you decide to take part you will be asked to complete and return a brief questionnaire in advance of a face-to-face interview with the researcher. These will involve answering some questions about processes for marking dissertations and discussion of work you identify as evidencing critical thinking. The interview should take no more than 45 minutes of your time. The process may (if necessary) with your permission be repeated after an appropriate period of time, anticipated to be six months to a year.

What are the possible disadvantages of taking part?

The researcher does not consider there to be any disadvantages of taking part as your responses will be anonymous and confidential. You are not being judged. You can choose whether or not to take part in this study.

What are the possible benefits of taking part?

The researcher hopes this study will aid understanding of critical thinking conceptually and as applied in assessment practices.

Will my taking part in this study be kept confidential?

The information that you provide will be anonymous and treated confidentially at all stages of the research and dissemination process. You will receive a copy of the interview transcription so you can check this for accuracy.

Data storage and access

All data will be stored securely in electronic and hard-copy format in compliance with the requirements of data protection legislation and associated Durham University School of Education regulations. A coding scheme will be employed to maintain your anonymity; it will not be possible to identify you directly or indirectly. No-one except the researcher and his academic supervisors will have access to the data. All members of the research team are bound by a strict code of ethics as academics of Durham University.

Use of the data

The anonymous data (which will not contain any personally identifiable information) collected for this study will form a key component of the researcher's doctoral thesis. It will further form the basis for presentations at academic conferences and subsequent journal publications.

Data retention

Data will be held securely and subsequently disposed of in accordance with Durham University School of Education policy.

Who is organising and funding the research?

The research is being organised and funded by Durham University Business School as part of a doctoral research programme. Gavin D'Northwood, Associate Professor (Teaching) in Accounting and part-time doctoral researcher will be undertaking the research. Gavin is supported by his academic Supervisors, Dr. Julie Rattray and Dr. Jonathan Tummons of Durham University School of Education.

Expenses and payments

The School's ethics policy does not allow any payment to be made to research participants.

Who has reviewed this study?

This research has received ethical approval from Durham University School of Education Research Ethics Committee.

Further information and contact details

If you would like any additional information about this project or have questions about the purpose of the research or how it will be carried out, please contact Gavin D'Northwood, the Principal Researcher (contact details supplied).

Appendix C: Pre-interview questionnaire (Pilot 1/Main)

Research Interview Pre-Questionnaire

Study title: **Conceptualising and operationalising assessment of critical thinking in Higher Education**

Researcher: Gavin D'Northwood, Associate Professor (Teaching) in Accounting, Durham University Business School.

Question 1

Evidence of 'critical thinking' is generally required for higher marks on dissertations. Can you describe what this means to you?

Question 2

Thinking about the most recent batch of dissertations that you have marked, can you tell me what you were looking for as evidence of critical thinking (up to 5 points)?

Can you think of a specific example of where you have seen ‘critical thinking’ in a dissertation? Can you identify how critical thinking was shown in that example?

Question 4

How do you think you have come to know what to look for when assessing critical thinking in dissertations? Do you feel your knowledge and understanding have developed over time?

Participant name:……. Participant signature:……

Date:……...

Appendix D: Pre-interview questionnaire (Pilot 2)

Research Interview Pre-Questionnaire

Study title: **Conceptualising and operationalising assessment of critical thinking in Higher Education**

Researcher: Gavin D'Northwood, Senior Teaching Fellow in Accounting,
Durham University Business School.

Question 1

Evidence of critical thinking is generally required for higher marks on dissertations.
Can you describe what this means to you?

Question 2

How do you think you first came to understand what is meant by critical thinking in the context of higher education?

Question 3

Thinking about the most recent batch of dissertations that you marked, can you explain what you were looking for as evidence of critical thinking (up to 5 points)?

Question 4

Can you think of a specific example of where you have seen critical thinking in a dissertation? Can you identify how critical thinking was shown in that example?

Question 5

How do you think you have come to know what to look for when assessing critical thinking in dissertations? What would you regard as instrumental in shaping this from your earliest recollection through to now?

Participant name:……

Participant signature:……

Date:…….....

Appendix E: Semi-structured interview guide (Pilot 2)

Research Interview Guide

Introduction

Thank you for agreeing to take part in this research project, in which I am investigating how critical thinking is understood as a concept by higher education tutors like yourself, to what extent that concept is shared amongst tutors and how tutors apply the concept when assessing dissertations. As indicated when I approached you to set up this interview I will be recording this discussion for later transcription and analysis. I trust this is ok with you?

Just to reiterate that your continuing participation is entirely voluntary and you are therefore free to withdraw at any time. Are we ok to continue?

I'm going to start with some questions on the dissertation marking process before re-visiting your answers to the pre-interview questionnaire, which you kindly completed prior to this interview, and finally then consider two examples from your most recent batch of dissertation marking.

Initial Questions

Q1. As a starting point, could you give me a brief history of your experience of teaching and assessing in your particular specialism?

Q2. Thinking of dissertations in particular, can you describe the process of marking and explain how you go about it?

Q3. What factors are important to you when marking dissertations?

Q4. What for you differentiates dissertations at different bands or grading levels?

Q5. How for you do written assessment criteria fit into the process of marking dissertations?

Artefact 1 – Pre-interview questionnaire

Discussion follows of the pre-interview questionnaire (Critical Thinking and assessment) – specifically, any questions and/or clarifications I might have following review prior to the interview. The questions from the pre-interview questionnaire are reprised here:

Q5. Evidence of critical thinking is generally required for higher marks on dissertations. Can you describe what this means to you?

Q6. Thinking about the most recent batch of dissertations that you have marked, can you explain what you were looking for as evidence of critical thinking (up to 5 points)?

Q7. Can you think of a specific example of where you have seen critical thinking in a dissertation? Can you identify how critical thinking was shown in that example?

Q8. How do you think you have come to know what to look for when assessing critical thinking in dissertations? Do you feel your knowledge and understanding have developed over time?

Artefacts 2 and 3 – 2 x example dissertations

Prior to this interview you have been engaged in marking dissertations and were asked to separate out two examples, one which rates highly for critical thinking and one which rates low or at least less so. For each of these, can you describe the differing levels of critical thinking as evidenced and explain how this has influenced your marks?

Appendix F: Data analysis coding lists

Coding - 1st round

Headings	Codes	Number of interviews tagged
Critical thinking - importance		**4**
Facets of critical thinking		
	Facet – argument	**18**
	Facet – big picture (able to see)	**6**
	Facet – creativity	**12**
	Facet – critique	**9**
	Facet – independent-mindedness	**0**
	Facet – inquisitiveness	**5**
	Facet – knowledge	**1**
	Facet – logic and reasoning	**9**
	Facet – open-mindedness	**13**
	Facet – problem solving	**2**
	Facet - scepticism	**10**
Finding critical thinking in dissertations		
	Critical thinking proxies	**2**
	Dissertation (1) Introduction – context	**5**
	Dissertation (1) Introduction – negatives	**2**
	Dissertation (1) Introduction – positives	**7**
	Dissertation (1) Introduction – setting up the research	**16**
	Dissertation (2) Literature Review – negatives	**20**
	Dissertation (2) Literature Review – positives	**20**

Superordinate codes	Sub-codes	Number of interviews coded
	Dissertation (3) Methodology and methods – negatives	**9**
	Dissertation (3) Methodology and methods – positives	**18**
	Dissertation (3) Methodology and methods – Quants vs. Qualitative	**5**
	Dissertation (4) Findings analysis and discussion - negatives	**8**
	Dissertation (4) Findings analysis and discussion - positives	**18**
	Dissertation (5) Conclusions - negatives	**2**
	Dissertation (5) Conclusions – positives (contribution)	**14**
	Dissertation (5) Conclusions – positives (limitations)	**12**
	Dissertation (5) Conclusions – positives (reflection)	**4**
	Dissertation (5) Conclusions – positives (research rigour)	**4**
	Dissertation (5) Conclusions – positives (top and tail)	**11**
	Dissertation (6) Holistic points – negatives	**11**
	Dissertation (6) Holistic points - positives	**17**
	Dissertation – approach	**7**
	Dissertation – communication	**11**
	Dissertation – originality (no)	**11**
	Dissertation – originality (yes)	**6**
	Dissertation – structure	**16**
	Dissertation – students' own thinking	**12**

Superordinate codes	Sub-codes	Number of interviews coded
Grading		
	Criteria (and critical thinking)	**3**
	Criteria – negatives (feedback match)	**8**
	Criteria – negatives (not used)	**5**
	Criteria – negatives (tacit or internal)	**5**
	Criteria – positives	**11**
	Criteria – positives (used)	**17**
	Criteria (post hoc)	**5**
	Experience counts	**9**
	Holistic assessment	**7**
	Constructed mark	**6**
	Forming a mark	**3**
	Professional judgement	**11**
Learning what critical thinking is		
	Learning through research	**8**
	Learning through subject immersion	**2**
	Learning through studying for qualifications	**3**
	Learning through supervising and marking	**6**
	Learning through working with academic colleagues (academic communities)	**7**
	Learning through working with academic colleagues (joint marking conversations)	**9**
	Learning through working with academic colleagues (mentoring conversations)	**3**
Making tacit knowledge (critical thinking in dissertations) explicit		
	Barriers	**1**
	Tacit to explicit (supervision)	**5**

Appendix G: Data analysis coding lists

Coding – 2nd round

Headings	Sub-headings	Codes	Number of interviews tagged
Category 1 – critical thinking experienced as an aspect of assessment processes			
	Grading		
		Criteria – link to critical thinking	**2**
		Criteria – negatives (feedback match)	**8**
		Criteria – negatives (not used)	**5**
		Criteria – negatives (post hoc check)	**7**
		Criteria – negatives (restrictions)	**5**
		Criteria – negatives (subjectivity)	**7**
		Criteria – negatives (tacit or internal)	**5**
		Criteria – positives (clarity)	**5**
		Criteria – positives (consistency)	**7**
		Criteria – positives (objectivity)	**7**
		Criteria – positives (used)	**16**
		Holistic assessment	**7**
		Constructed mark	**6**
		Forming a mark	**3**
		Professional judgement	**11**
	Making tacit knowledge (critical thinking in dissertations) explicit		
		Barriers	**1**
		Tacit to explicit (supervision)	**7**

Headings	Sub-headings	Codes	Number of interviews tagged
Category 2 – critical thinking experienced as facets of a critical thinker			
	Facets of a critical thinker		
		Facet – argument	**18**
		Facet – big picture (able to see)	**6**
		Facet – creativity	**12**
		Facet – critique	**9**
		Facet – inquisitiveness	**5**
		Facet – logic and reasoning	**9**
		Facet – open-mindedness	**14**
		Facet - scepticism	**12**
Category 3 – critical thinking experienced as an aspect of undergraduate dissertations			
	Finding critical thinking in undergraduate dissertations		
	Dissertation (1) - Introduction		
		Setting up the research	**16**
		Contextual base	**6**
		Focus – aims, objectives	**14**
		Interest and motivation	**12**
		Literature base	**7**
		Originality – no	**11**
		Originality – yes	**6**
	Dissertation (2) – Literature review		
		Negative - description	**16**
		Negative – focus	**2**
		Negative – references	**5**
		Negative – selectivity	**2**
		Positive – compare and contrast	**16**
		Positive – conceptualisation	**4**

Headings	Sub-headings	Codes	Number of interviews tagged
		Positive – critique	3
		Positive – depth	5
		Positive – evaluation	7
		Positive – identify the gap	8
		Positive – references	10
		Positive – RQ-LR-Methodological fit)	14
		Positive – scope	11
		Positive – synthesis	7
		Positive – thematic	6
		Positive – theorisation	10
	Dissertation (3) – Methodology		
		Negative – generic	2
		Negative – replication studies	5
		Negative – research rigour	3
		Negative – RQ-LR-Methodological fit	6
		Positive – flexibility	5
		Positive – replication studies	3
		Positive – research rigour	9
		Positive – RQ-LR-Methodological fit (alignment)	10
		Positive – RQ-LR-Methodological fit (justification)	13
	Dissertation (4) Findings analysis and discussion		
		Negative – analysis & interpretation	9
		Negative – description	2
		Negative – evidence	2

Headings	Sub-headings	Codes	Number of interviews tagged
		Negative – RQ-LR-Methodological-F&A fit	3
		Negative - synthesis (LR)	2
		Positive – analysis & interpretation	13
		Positive – insight	3
		Positive – theorisation	6
		Positive – evaluation	2
		Positive - evidence	3
		Positive – RQ-LR-Methodological-F&A fit	4
		Positive – synthesis (LR)	15
	Dissertation (5) Conclusions		
		Contribution	15
		Limitations	13
		Reflection	4
		Research rigour	4
		Top & tail	12
	Dissertation (6) Holistic points		
		Consistency	16
		Negative – basic	6
		Negative – consistency	9
		Positive – consistency	13
		Positive - depth	2
	Dissertation (7) Presentational aspects		
		Communication	11
		Structure	16
		Students' own thinking	12

Headings	Sub-headings	Codes	Number of interviews tagged
Category 4 – Critical thinking experienced as situated learning			
	How participants have learned what critical thinking is		
		Learning through research	**9**
		Learning through subject immersion	**5**
		Learning through supervising and marking	**6**
		Learning through peer interactions (institutional contexts)	**3**
		Learning through peer interactions (marking conversations)	**14**
		Learning through peer interactions (mentoring conversations)	**3**
		Learning through peer interactions (sharing & collaboration)	**2**
		Learning through peer interactions (wider engagement)	**4**

Appendix H: Data analysis coding lists

Coding – 3rd round

Headings	Sub-headings	Codes	Number of interviews tagged
Category 1 – critical thinking experienced as an aspect of assessment processes			
	Grading		
		Criteria – link to critical thinking	**2**
		Criteria – negatives (feedback match)	**8**
		Criteria – negatives (not used)	**5**
		Criteria – negatives (post hoc check)	**6**
		Criteria – negatives (restrictions)	**5**
		Criteria – negatives (subjectivity)	**7**
		Criteria – negatives (tacit or internal)	**5**
		Criteria – positives (clarity)	**5**
		Criteria – positives (consistency)	**6**
		Criteria – positives (objectivity)	**6**
		Criteria – positives (used)	**16**
		Holistic assessment	**7**
		Constructed mark	**6**
		Forming a mark	**3**
		Professional judgement	**12**
	Making tacit knowledge (critical thinking in dissertations) explicit		
		Barriers	**1**
		Tacit to explicit (supervision)	**7**

Headings	Sub-headings	Codes	Number of interviews tagged
Category 2 – critical thinking experienced as facets of a critical thinker			
	Facets of a critical thinker		
		Facet – argument	**21**
		Facet – big picture (able to see)	**6**
		Facet – creativity	**12**
		Facet – critique	**9**
		Facet – inquisitiveness	**5**
		Facet – logic and reasoning	**9**
		Facet – open-mindedness	**15**
		Facet - scepticism	**12**
Category 3 – critical thinking experienced as an aspect of undergraduate dissertations			
	Finding critical thinking in undergraduate dissertations		
	Dissertation (1) - Introduction		
		Setting up the research	**16**
		Contextual base	**7**
		Focus – aims, objectives	**17**
		Interest and motivation	**10**
		Justification	**8**
		Literature base	**9**
		Originality – no	**12**
		Originality – yes	**6**
	Dissertation (2) – Literature review		
		Negative - description	**17**
		Negative – focus	**2**
		Negative – references	**4**
		Negative – selectivity	**2**
		Positive – compare and contrast	**17**

Headings	Sub-headings	Codes	Number of interviews tagged
		Positive – conceptualisation	5
		Positive – critique	3
		Positive – depth	5
		Positive – evaluation	7
		Positive – identify the gap	8
		Positive – references	9
		Positive – RQ-LR-Methodological fit)	14
		Positive – scope	11
		Positive – synthesis	8
		Positive – thematic	7
		Positive – theorisation	11
	Dissertation (3) – Methodology		
		Negative – feasibility	1
		Negative – generic	2
		Negative – replication studies	5
		Negative – research rigour	3
		Negative – RQ-LR-Methodological fit	5
		Positive – flexibility	3
		Positive - references	3
		Positive – replication studies	2
		Positive – research rigour	7
		Positive – RQ-LR-Methodological fit (alignment)	10
		Positive – RQ-LR-Methodological fit (justification)	13
	Dissertation (4) Findings analysis and discussion		
		Negative – analysis & interpretation	8

Headings	Sub-headings	Codes	Number of interviews tagged
		Negative – description	3
		Negative – evidence	2
		Negative – RQ-LR-Methodological-F&A fit	3
		Negative - synthesis (LR)	2
		Positive – analyse	10
		Positive – insight	4
		Positive – perspectives	2
		Positive – theorisation	7
		Positive – evaluation	4
		Positive - evidence	4
		Positive – RQ-LR-Methodological-F&A fit	4
		Positive – synthesis (LR)	16
	Dissertation (5) Conclusions		
		Contribution	15
		Limitations	13
		Reflection	5
		Research rigour	4
		Top & tail	12
	Dissertation (6) Holistic points		
		Negative – basic	5
		Negative – consistency	9
		Positive – consistency	13
		Positive – depth	2
	Dissertation (7) Presentational aspects		
		Communication	11
		Structure	16
		Students' own thinking	13

Headings	Sub-headings	Codes	Number of interviews tagged
Category 4 – Critical thinking experienced as situated learning			
	How participants have learned what critical thinking is		
		Learning through research	**11**
		Learning through subject immersion	**6**
		Learning through supervising and marking	**9**
		Learning through peer interactions (institutional contexts)	**3**
		Learning through peer interactions (marking conversations)	**14**
		Learning through peer interactions (mentoring conversations)	**3**
		Learning through peer interactions (sharing & collaboration)	**3**
		Learning through peer interactions (wider engagement)	**5**

Appendix I: Data analysis coding lists

Coding – 4th round

Headings	Codes	Number of interviews tagged
(1) Critical thinking as dispositions		
	Inquisitiveness	**5**
	Scepticism	**12**
	Open-mindedness	**14**
	Big picture	**5**
	Diss 1 - Intro - justification, interest & motivation	**12**
	Diss 1 - Intro - focus, aims, objectives & RQs	**14**
	Diss 3 - M&Ms - flexibility	**5**
	Diss 5 - Conc - limitations	**13**
(2) Critical thinking as argument		
	Argument	**18**
	Logic and reasoning	**9**
	Diss 4 - F&A – evidence	**4**
	Diss 1 - Intro - contextual base	**6**
	Diss 1 - Intro - literature base	**8**
	RQ-LR-M&M fit	**18**
	Research rigour	**10**
	Structure	**21**
	Communication	**11**
(3) Critical thinking as judgement		
	Critique	**12**
	Analysis & interpretation	**16**
	Diss 2 - LR – focus	**16**
	Diss 2 - LR - compare and contrast	**16**
	Diss 2 - LR - evaluation	**7**
	Diss 2 - LR - identify the gap	**8**

Headings	Codes	Number of interviews tagged
	Synthesis	9
	Students' own thinking	12
(4) Critical thinking as originality		
	Originality (no)	11
	Originality (yes)	5
	Contribution	15
	Creativity	12
	Diss 1 - Intro - focus, aims, objectives & RQs	13
	Description/basic	18
	Replication studies	6
(5) Critical thinking assessment as socially situated		
	Professional judgement in assessing	11
	Holistic vs constructed mark	6
	Learning through assessment structures (criteria -ve)	16
	Learning through assessment structures (criteria +ve)	17
	Learning through research	9
	Learning through subject immersion	5
	Learning through supervising students	6
	Learning from peers	16

Appendix J: Data analysis coding lists

Coding – 5th round

Superordinate codes	Sub-codes	Number of interviews coded
(1) Critical thinking as dispositions		
	Inquisitiveness	**13**
	Scepticism	**18**
	Open-mindedness	**15**
	Big picture	**5**
(2) Critical thinking as skills/abilities		
	Logic and reasoning	**9**
	Critique	**19**
	Analysis & interpretation	**19**
	Synthesis	**9**
	Argument (key skill)	**18**
	Argument (evidence)	**13**
	Argument (structure)	**21**
	Argument (RQ-LR-M&M fit)	**18**
	Argument (communication clarity)	**10**
	Argument (research rigour)	**10**
(3) Critical thinking as originality		
	Originality required (yes)	**5**
	Originality required (no)	**11**
	Originality is students' own thinking	**12**
	Originality is contribution	**15**
	Originality is creativity	**12**
	Originality is not description/basic	**18**
	Originality is not replication studies	**6**
	Originality is not safety first approach	**9**

Superordinate codes	Sub-codes	Number of interviews coded
(4) Critical thinking learned		
	Through research and publishing	**9**
	Through subject immersion	**5**
	Through supervising students	**6**
	Through interaction with peers	**16**
(5) Approaches to assessment: similarities and differences		
	Professional judgement	**11**
	Holistic approach	**9**
	Constructed approach	**6**
	Assessment criteria (used when assessing)	**16**
	Assessment criteria (used for feedback)	**8**
	Assessment criteria (used as post hoc check)	**7**
	Assessment criteria (tacit/internal criteria used)	**5**
	Assessment criteria (not used)	**5**
	Originality required (yes)	**5**
	Originality required (no)	**11**

References

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 91-106. https://doi.org/10.1080/0969594X.2011.650150

Alderman, G. (2009). Defining and measuring academic standards: A British perspective. *Higher Education Management and Policy*, *21*(3), 1-14. https://doi.org/10.1787/17269822

Amabile, T. M. 1996. Creativity in Context: Update to 'The Social Psychology of Creativity'. Oxford: Westview Press.

Anastasiadou, S., & Dimitriadou, A. (2011). What does Critical Thinking mean? A statistical data analysis of pre-service teachers' defining statements. *International Journal of Humanities and Social Science*, *1*(7), 73-83. Retrieved July 18, 2019, from http://www.ijhssnet.com/view.php?u=http://www.ijhssnet.com/journals/Vol._1_No._7_[Special_Issue_June_2011]/11.pdf

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research, and Evaluation*, *10*(1), 3. Retrieved March 29, 2016 from https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1152&context=pare

Andrews, R. (1995). *About argument: Teaching and learning argument*. London: Continuum International Publishing Group Ltd.

Andrews, R. (2003). The end of the essay?. *Teaching in Higher Education*, *8*(1), 117-128. https://doi.org/10.1080/1356251032000052366

Andrews R. (2015). Critical Thinking and/or Argumentation in Higher Education. In: M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education,* (pp. 49-62). Palgrave Macmillan, New York. https://doi.org/10.1057/9781137378057_3

Andrews, R. (2015). Critical thinking and/or argumentation in higher education. In *The Palgrave handbook of critical thinking in higher education* (pp. 49-62). New York: Palgrave Macmillan.

Andrews, T. (2012). What is social constructionism?. *Grounded theory review*, *11*(1). Retrieved July 3, 2019, from https://web-b-ebscohost-com.ezphost.dur.ac.uk/ehost/pdfviewer/pdfviewer?vid=5&sid=b45075b2-cdf1-4488-93fe-962e3ef2d30d%40sessionmgr101

Arum, R., & Roksa, J. (2008). Learning to Reason and Communicate in College: Initial Report of Findings from the CLA Longitudinal Study. *Social Science Research Council*. Retrieved March 22, 2016, from https://files.eric.ed.gov/fulltext/ED514992.pdf

Ashwin, P., Abbas, A., & McLean, M. (2017). How does completing a dissertation transform undergraduate students' understandings of disciplinary knowledge?. *Assessment & Evaluation in Higher Education, 42*(4), 517-530. https://doi.org/10.1080/02602938.2016.1154501

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly, 31*(1), 71–94. https://doi.org/10.2307/3587975

Bailin, S., & Siegel, H. (2002). Critical thinking. In: N. Blake, P. Smeyers, R. Smith & P. Standish (Eds.), *The Blackwell guide to the philosophy of education*. (pp. 181-193). Oxford: Blackwell Publishing Ltd.

Bailin, S., Case, R., Coombs, J. & Daniels, L. (1999a). Common misconceptions of critical thinking. *Journal of Curriculum Studies*, 31 (3), 269–283. https://doi.org/10.1080/002202799183124

Bailin, S., Case, R., Coombs, J. & Daniels, L. (1999b). Conceptualising critical thinking. *Journal of Curriculum Studies*, 31 (3), 285–302. https://doi.org/10.1080/002202799183133

Baird, J. A., Greatorex, J., & Bell, J. F. (2004). What makes marking reliable? Experiments with UK examinations. *Assessment in education: Principles, policy & practice*, *11*(3), 331-348. https://doi.org/10.1080/0969594042000304627

Bamber, M. (2015). The impact on stakeholder confidence of increased transparency in the examination assessment process. *Assessment & Evaluation in Higher Education*, *40*(4), 471-487. http://dx.doi.org/10.1080/02602938.2014.921662

Baptista, A., Frick, L., Holley, K., Remmik, M., & Tesch, J. (2015). The Doctorate as an Original Contribution to Knowledge: Considering Relationships between Originality, Creativity, and Innovation. *Frontline Learning Research*, *3*(3), 55-67. Retrieved August 13, 2020 from https://files.eric.ed.gov/fulltext/EJ1091022.pdf

Baril, C. P., Cunningham, B. M., Fordham, D. R., Gardner, R. L., & Wolcott, S. K. (1998). Critical thinking in the public accounting profession: Aptitudes and attitudes. *Journal of Accounting Education*, *16*(3-4), 381-406. https://doi.org/10.1016/S0748-5751(98)00023-2

Barkaoui, K. (2010). Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study. *Tesol Quarterly*, *44*(1), 31-57. https://doi.org/10.5054/tq.2010.214047

Barnacle, R. (2005). Interpreting interpretation: A phenomenological perspective on phenomenography. In: J. A. Bowden & P. Green. (Eds.), *Doing developmental phenomenography*. (pp. 47–55). Melbourne: RMIT University Press.

Barnet, S., & H. Bedau. (2010). *Critical Thinking, Reading, and Writing: A Brief Guide to Argument* (7th ed.). Bedford: Boston, MA.

Barnett, R. (1997). *Higher education: A critical business*. Milton Keynes: Open University Press.

Barrie, S., & Prosser, M. (2004). Generic graduate attributes: Citizens for an uncertain future. *Higher Education Research and Development, 23*(3), 243-246. https://doi.org/10.1080/0729436042000235373

Bassey, M. (1999). *Case study research in educational settings*. Buckingham: Open University Press.

BATTERSBY, M. E. (1989). Critical Thinking as Applied Epistemology: Relocating Critical Thinking in the Philosophical Landscape. *Informal Logic, 11*(2), 91-100. Retrieved August 5, 2020, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.894.5514&rep=rep1&type=pdf

Battersby, M. (2018). Critical thinking as applied epistemology. In M. Battersby & S. Bailin (Eds.), *Inquiry: A New Paradigm for Critical Thinking, 7*. (pp. 2-20). Windsor, Ontario: Windsor Studies in Argumentation.

Baume, D. S., Yorke, M., & Coffey, M. (2004). What is happening when we assess, and how can we use our understanding of this to improve assessment?. *Assessment & Evaluation in Higher Education, 29*(4), 451-477. http://dx.doi.org/10.1080/02602930310001689037

Baxter Magolda, M. B. (1992) *Knowing and Reasoning in College Students; gender-related patterns in students' intellectual development*. San Francisco, Jossey-Bass

Baxter Magolda, M. B. (2002). Helping students make their way to adulthood: Good company for the journey. *About Campus*, *6*(6), 2-9. https://doi.org/10.1177%2F108648220200600602

Bazeley, P., & Jackson, K. (Eds.). (2013). Qualitative data analysis with NVivo. London: SAGE Publications Ltd.

Beghetto, R. A. (2013). *Killing ideas softly?: The promise and perils of creativity in the classroom*. Charlotte, NC: Information Age Publishing, Inc.

Bell, J. (2014). *Doing Your Research Project: A guide for first-time researchers* (6th ed.) Maidenhead: McGraw-Hill Education (UK).

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS quarterly*, 369-386. Retrieved July 3, 2019, from https://www.jstor.org/stable/248684?seq=1

Bennich-Björkman, L. 1997. Organising Innovative Research: The Inner Life of University Departments. Oxford: Pergamon IAU Press.

Bensley, D. A., Rainey, C., Murtagh, M. P., Flinn, J. A., Maschiocchi, C., Bernhardt, P. C., & Kuehne, S. (2016). Closing the assessment loop on critical thinking: The challenges of multidimensional testing and low test-taking motivation. *Thinking Skills and Creativity*, *21*, 158-168. https://doi.org/10.1016/j.tsc.2016.06.006

Berger, P. & Luckmann, T. (1991). *The social construction of reality*. London: Penguin Books.

Berheide, C. W. (2007). Doing Less Work, Collecting Better Data: Using Capstone Courses to Assess Learning. Peer Review, 9(2), 27-30. Retrieved December 23, 2019, from https://www.researchgate.net/profile/Catherine_Berheide/publication/270891171_Doing_Less_Work_Collecting_Better_Data_Using_Capstone_Courses_to_Assess_Learning/links/574dd9e108ae8bc5d15bf74e.pdf

Bernstein, D. A. (1995). A negotiation model for teaching critical thinking. *Teaching of Psychology, 22*(1), 22-24.

Bettany-Saltikov, J., Kilinc, S., & Stow, K. (2009). Bones, boys, bombs and booze: an exploratory study of the reliability of marking dissertations across disciplines. *Assessment & Evaluation in Higher Education*, *34*(6), 621-639. https://doi.org/10.1080/02602930802302196

Beutel, D., Adie, L., & Lloyd, M. (2017). Assessment moderation in an Australian context: processes, practices, and challenges. *Teaching in Higher Education, 22*(1), 1-14. https://doi.org/10.1080/13562517.2016.1213232

Bird, C. M. (2005). How I stopped dreading and learned to love transcription. *Qualitative inquiry*, *11*(2), 226-248. https://doi.org/10.1177/1077800404273413

Bird, F. L., & Yucel, R. (2013). Improving marking reliability of scientific writing with the Developing Understanding of Assessment for Learning programme. *Assessment & Evaluation in Higher Education, 38*(5), 536-553. http://dx.doi.org/10.1080/02602938.2012.658155

Bissell, A. N., & Lemons, P. P. (2006). A New Method for Assessing Critical Thinking in the Classroom. *BioScience*, *56*(1), 66–72. https://doi.org/10.1641/0006-3568(2006)056[0066:ANMFAC]2.0.CO;2

Black, B. (2012). An overview of a programme of research to support the assessment of Critical Thinking. *Thinking Skills and Creativity*, *7*(2), 122-133. https://doi.org/10.1016/j.tsc.2012.04.003

Bloxham, S. (2009). Marking and moderation in the UK: false assumptions and wasted resources. *Assessment & Evaluation in Higher Education, 34*(2). http://dx.doi.org/10.1080/02602930801955978

Bloxham, S., & Boyd, P. (2012). Accountability in grading student work: securing academic standards in a twenty-first century quality assurance context. *British Educational Research Journal, 38*(4), 615-634. http://dx.doi.org/10.1080/01411926.2011.569007

Bloxham, S., & Price, M. (2015). External examining: fit for purpose?. *Studies in Higher Education*, *40*(2), 195-211. https://doi.org/10.1080/03075079.2013.823931

Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education, 36*(6), 655-670. http://dx.doi.org/10.1080/03075071003777716

Bloxham, S., Hudson, J., den Outer, B., & Price, M. (2015). External peer review of assessment: An effective approach to verifying standards?. *Higher Education Research & Development*, *34*(6), 1069-1082. http://dx.doi.org/10.1080/07294360.2015.1024629

Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016a). Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education, 41*(3), 466-481. http://dx.doi.org/10.1080/02602938.2015.1024607

Bloxham, S., Hughes, C., & Adie, L. (2016b). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment & Evaluation in Higher Education, 41*(4), 638-653. http://dx.doi.org/10.1080/02602938.2015.1039932

Bok, D. (2006). *Our underachieving colleges. A candid look at how much students learn and why they should be learning more.* Princeton, NJ: Princeton University Press.

Boud, D. (2007). Reframing assessment as if learning was important. In: D. Boud, & N. Falchikov (Eds.), *Rethinking Assessment for Higher Education: Learning for the Longer Term* (pp. 14-25). London: Routledge.

Brabeck, M. M. (1983). Critical thinking skills and reflective judgment development: Redefining the aims of higher education. *Journal of Applied Developmental Psychology*, *4*(1), 23-34. https://doi.org/10.1016/0193-3973(83)90056-4

Brannick, T., & Coghlan, D. (2007). In defense of being "native": The case for insider academic research. *Organizational research methods*, *10*(1), 59-74. https://doi.org/10.1177/1094428106289253

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V. & Clarke, V. (2012) Thematic analysis. In: H. Cooper, H. (Ed.), *The Handbook of Research Methods in Psychology*. Washington, DC: American Psychological Association

Brennan, J., & Osborne, M. (2005). The organisational mediation of university learning. Working Paper. Higher Education Academy and Open University/CHERI, York. Retrieved May 19, 2020, from https://eprints.gla.ac.uk/117383/

Brennan, N. (1998). *Accounting research: a practical guide*. Dublin: Oak Tree Press.

Brinkmann, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Brodin, E. M. (2016). Critical and creative thinking nexus: learning experiences of doctoral students. *Studies in Higher Education*, *41*(6), 971-989. https://doi.org/10.1080/03075079.2014.943656

Brookfield, S. D. (1987). *Developing critical thinkers: Challenging adults to explore alternative ways of thinking and acting*. San Francisco: Jossey-Bass.

Brooks, V. (2012). Marking as judgment. *Research Papers in Education, 27*(1), 63-80. https://doi.org/10.1080/02671520903331008

Brown, K., & Rutter, L. (2008). *Critical thinking for Social Work* (2nd ed.). Exeter: Learning Matters Ltd.

Browne, M.N., & Freeman, K. (2000). Distinguishing Features of Critical Thinking Classrooms. *Teaching in Higher Education, 5*(3), 301-309. https://doi.org/10.1080/713699143

Brownlee, J. (2004). Teacher education students' epistemological beliefs: Developing a relational model of teaching. *Research in Education*, *72*(1), 1-17. Retrieved April 10, 2018, from https://doi.org/10.7227/RIE.72.1

Bryman, A. (2015). *Social Research Methods* (5th ed.). Oxford: Oxford University Press.

Bryman, A. & Bell, E. (2011). *Business Research Methods* (3rd ed.). Oxford: Oxford University Press.

Bryman, A., & Burgess, R. G. (1994). Reflections on qualitative data analysis. In: A. Bryman & R. G. Burgess (Eds.), *Analyzing qualitative data*, 216-226. London: Routledge.

Burr, V. (2003). *Social Constructionism* (2nd ed.). London: Routledge.

Buskist, W., & Irons, J. G. (2008). Simple strategies for teaching your students to think critically. In: D. S. Dunn, J. S. Halonen, & R. A. Smith (Eds.), *Teaching Critical Thinking in Psychology: A Handbook of Best Practices* (pp. 49-57). London: Wiley & Blackwell.

Buswell, N. T. (2018, October). Behind the scenes: Course syllabi explained. In *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE. Retrieved April 8, 2021, from https://ieeexplore.ieee.org/abstract/document/8659273

Buswell, N. T., & Berdanier, C. G. (2020, October). Revealing teaching conceptions and methods through document elicitation of course syllabi and statements of teaching philosophy. In *2020 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE. Retrieved April 8, 2021, from https://aic-atlas.s3.eu-north-1.amazonaws.com/projects/e7299991-eb2b-4764-a849-4909e01fb07d/documents/G6ONkqZ4tggHKDJqUwvskyDi5dvYZK231QkDobCJ.pdf

Butler, H. A., & Halpern, D. F. (2020). Critical thinking impacts our everyday lives. In Sternberg R. J., & Halpern, D. F. (Eds.), *Critical Thinking in Psychology*. (pp. 152–162). Cambridge: Cambridge University Press.

Calma, A., & Davies, M. (2020). Critical thinking in business education: current outlook and future prospects. *Studies in Higher Education*, 1-17. https://doi.org/10.1080/03075079.2020.1716324

Cannings, R., Hawthorne, K., Hood, K., & Houston, H. (2005). Putting Double Marking to the Test: A Framework to Assess If It is worth the Trouble. *Medical Education, 39*(3), 299–308. https://doi.org/10.1111/j.1365-2929.2005.02093.x

Carrington, M., Chen, R., Davies, M., Kaur, J., & Neville, B. (2011). The effectiveness of a single intervention of computer-aided argument mapping in a marketing and a financial accounting subject. *Higher Education Research & Development, 30*(3), 387-403. http://dx.doi.org/10.1080/07294360.2011.559197

Carrithers, D., & Bean, J. C. (2008). Using a client memo to assess critical thinking of finance majors. *Business Communication Quarterly*, *71*(1), 10-26. https://doi.org/10.1177%2F1080569907312859

Carroll, D. W. (2007). Patterns of student writing in a critical thinking course: A quantitative analysis. *Assessing writing*, *12*(3), 213-227. https://doi.org/10.1016/j.asw.2008.02.001

Chen, D. L., & Rattray, J. (2017). Transforming thinking through problem-based learning in the news media literacy class: Critical thinking as a threshold concept towards threshold capabilities. *Practice and Evidence of the Scholarship of Teaching and Learning in Higher Education*, *12*(2), 272-293.

Clarke, G., & Lunt, I. (2014). The concept of 'originality' in the Ph. D.: how is it interpreted by examiners?. *Assessment & evaluation in higher education*, *39*(7), 803-820. https://doi.org/10.1080/02602938.2013.870970

Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The psychologist*, *26*(2), 120-133. Retrieved July 9, 2019, from https://uwe-repository.worktribe.com/output/937596/teaching-thematic-analysis-overcoming-challenges-and-developing-strategies-for-effective-learning

Clarke, V. & Braun, V. (2014) Thematic analysis. In A. C. Michalos (Ed.), *Encyclopaedia of Quality of Life and Well-Being Research* (pp. 6626-6628). Dordrecht, Netherlands: Springer.

Clarke, V., & Braun, V. (2017). Thematic analysis. The Journal of Positive Psychology, 12(3), 297-298. http://dx.doi.org/10.1080/17439760.2016.1262613

Clarke, V., & Braun, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Counselling and Psychotherapy Research*, *18*(2), 107-110. https://doi.org/10.1002/capr.12165

Clegg, K., & Bryan, C. (2006). Reflections, rationales and realities. In: C. Bryan, & K. Clegg (Eds.), *Innovative assessment in higher education*. (pp. 216-227). Abingdon: Routledge.

Cohen, L., Duberley, J., & Mallon, M. (2004). Social constructionism in the study of career: Accessing the parts that other approaches cannot reach. *Journal of Vocational Behavior*, *64*(3), 407-422. https://doi.org/10.1016/j.jvb.2003.12.007

Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education*. (8th ed.). London: Routledge.

Colley, H., & Silver, H. (2005). External examiners and the benchmarking of standards. *Higher Education Academy, York*.

Collins English Dictionary Online (2020). Retrieved 23 December 2020 from https://www.collinsdictionary.com/dictionary/english

Collins English Thesaurus Online (2020). Retrieved 7 April 2020 from https://www.collinsdictionary.com/dictionary/english

Coolican, H., (2009). *Research Methods and Statistics in Psychology* (5th ed.). Oxon: Routledge.

Cooper, D. J., & Morgan, W. (2008). Case study research in accounting. *Accounting horizons*, *22*(2), 159-178. http://doi.org/10.2308/acch.2008.22.2.159

Cooper, S., & Patton, R. (2009). *Writing logically, thinking critically* (6th ed.). New York: Pearson.

Cosgrove, R. (2011). Critical thinking in the Oxford tutorial: a call for an explicit and systematic approach. *Higher Education Research & Development, 30*(3), 343-356. http://dx.doi.org/10.1080/07294360.2010.487259

Cottrell, S. (2017). *Critical thinking skills: Effective analysis, argument and reflection* (3rd ed.). London: Palgrave, Macmillan International Higher Education.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. London: SAGE Publications Ltd.

Cunliffe, A. L. (2008). Orientations to social constructionism: Relationally responsive social constructionism and its implications for knowledge and learning. *Management learning*, *39*(2), 123-139. https://doi.org/10.1177/1350507607087578

Dai, N. T., Free, C., & Gendron, Y. (2019). Interview-based research in accounting 2000–2014: Informal norms, translation and vibrancy. *Management Accounting Research*, *42*, 26-38. https://doi.org/10.1016/j.mar.2018.06.002

Dalziel, J. (1998). Using Marks to Assess Student Performance, some problems and alternatives. *Assessment & evaluation in higher education*, *23*(4), 351-366. https://doi.org/10.1080/0260293980230403

Danvers, E. (2018). Who is the critical thinker in higher education? A feminist re-thinking. *Teaching in Higher Education, 23*(5), 548-562. https://doi.org/10.1080/13562517.2018.1454419

Davies, W. M. (2006). An 'infusion' approach to critical thinking: Moore on the critical thinking debate. *Higher Education Research & Development*, *25*(2), 179-193. http://dx.doi.org/10.1080/07294360600610420

Davies, M. (2011). Introduction to the special issue on critical thinking in higher education. *Higher Education Research & Development*, *30*(3), 255-260. http://dx.doi.org/10.1080/07294360.2011.562145

Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development, 32*(4), 529-544. http://dx.doi.org/10.1080/07294360.2012.697878

Davies, M. (2015). A Model of Critical Thinking in Higher Education. In: M. Paulsen (Eds.), *Higher Education: Handbook of Theory and Research, 30*. (pp. 41-92). Springer, Cham. https://doi.org/10.1007/978-3-319-12835-1_2

Davos World Economic Forum. (2019). This soft skill will boost your employability. Retrieved July 3, 2020, from https://www.weforum.org/agenda/2019/10/critical-thinking-key-skills-future-work/

de Bono, E. (1982). *de Bono's Thinking Course.* London: British Broadcasting Corporation (BBC).

de Bono, E. (1983). *Practical Thinking.* Harmondsworth: Penguin

de la Fuente, J. R. (2009). UNESCO World Conference on Higher Education WCHE 2009. New dynamics of higher education and research for societal change and development. *Universidades*, 42, 3-5. Retrieved April 8, 2020 from https://www.redalyc.org/pdf/373/37313030002.pdf

Denzin, N. K., & Lincoln, Y. S. (2008). *Introduction: The discipline and practice of qualitative research.* In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (p. 1–43). Thousand Oaks, CA: Sage Publications, Inc.

Denzin, N. K., & Lincoln, Y. S. (2011). *Introduction: The discipline and practice of qualitative research.* In: N. K. Denzin & Y. S. Lincoln (Eds.). *The Sage handbook of qualitative research*, 4. (pp. 1-20). Thousand Oaks, CA: SAGE Publications, Inc.

Department for Business, Innovation and Skills (BIS). (2016). Success as a Knowledge Economy: teaching excellence, social mobility and student choice, HMSO, London. Retrieved May 19, 2020, from https://www.gov.uk/government/publications/higher-education-success-as-a-knowledge-economy-white-paper

Dortins, E. (2002). Reflections on phenomenographic process: Interview, transcription and analysis. *Quality conversations: Research and development in higher education*, *25*, 207-213.

Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral vision: Expertise in real world contexts. *Organization studies*, *26*(5), 779-792. https://doi.org/doi/pdf/10.1177/0170840605053102

du Boulay, D. (1999). Argument in Reading: what does it involve and how can students become better critical readers?, *Teaching in Higher Education, 4*(2), 147-162. http://dx.doi.org/10.1080/1356251990040201

Dumitru, D. (2019). Creating meaning. The importance of Arts, Humanities and Culture for critical thinking development. *Studies in Higher Education, 44*(5), 870-879. https://doi.org/10.1080/03075079.2019.1586345

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking skills and Creativity*, *12*, 43-52. https://doi.org/10.1016/j.tsc.2013.12.004

Ebel, R.L. (1965). *Measuring educational achievement*. New Jersey: Prentice Hall.

Ecclestone, K. (2001). 'I know a 2:1 when I see it': Understanding criteria for degree classifications in franchised university programmes. *Journal of Further and Higher Education, 25*(3), 301-313. http://dx.doi.org/10.1080/03098770126527

Eisner, E. W. (1985). *The art of educational evaluation a personal view*. London: Falmer Press.

Ekstrom, R. B., Villegas, A. M., Bruschi, B. A., & Cline, F. A. (1994). College grades: An exploratory study of policies and practices. *ETS Research Report Series*, *1994*(1), i-33. https://doi.org/10.1002/j.2333-8504.1994.tb01596.x

Elton, L., & Johnston, B. (2002). Assessment in Universities: a critical review of research. Retrieved March 29, 2016, from https://eprints.soton.ac.uk/59244/1/59244.pdf

Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*, *27*(3), 243-256. https://doi.org/10.1080/02602930220138606

Ennis, R. H. (1962). A concept of critical thinking. *Harvard educational review, 32*(1), 81-111.

Ennis, R. H. (1985a). A logical basis for measuring critical thinking skills. *Educational leadership*, *43*(2), 44-48. Retrieved April 12, 2018 from https://jgregorymcverry.com/readings/ennis1985assessingcriticalthinking.pdf

Ennis, R. H. (1985b). Critical thinking and the curriculum. *Thinking skills instruction: Concepts and techniques*, 40-48. Retrieved June 15, 2020, from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1006.6819&rep=rep1&type=pdf#page=41

Ennis, R. H. (1987). *A taxonomy of critical thinking dispositions and abilities.* In J. B. Baron & R. J. Sternberg (Eds.), *Series of books in psychology. Teaching thinking skills: Theory and practice* (p. 9-26). NY: W H Freeman/Times Books/ Henry Holt & Co.

Ennis, R. H. (1993). Critical Thinking Assessment. *Theory Into Practice, 32*(3), 179-186. https://doi.org/10.1080/00405849309543594

Ennis, R. H. (1996). Critical thinking dispositions: Their nature and Asses

sability. *Informal logic*, *18*(2). Retrieved June 15, 2020, from https://ojs.uwindsor.ca/index.php/informal_logic/article/view/2378

Erikson, M. G., & Erikson, M. (2019). Learning outcomes and critical thinking–good intentions in conflict. *Studies in Higher Education*, *44*(12), 2293-2303. https://doi.org/10.1080/03075079.2018.1486813

Espey, M. (2018). Enhancing critical thinking using team-based learning. *Higher Education Research & Development, 37*(1), 15-29. https://doi.org/10.1080/07294360.2017.1344196

Evans, N. J., Forney, D. S., Guido, F. M., Patton, L. D., & Renn, K. A. (2010). *Student development in college: Theory, research, and practice.* (2nd ed.). San Francisco: Jossey Bass.

Facione, P.A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*, 1-112. Retrieved December 23, 2016, from https://philarchive.org/archive/FACCTA

Facione, P. A. (2011). Critical thinking: What it is and why it counts. *Insight assessment*, *2007*(1), 1-23. Retrieved December 23, 2016, from https://www.researchgate.net/profile/Peter_Facione/publication/251303244_Critical_Thinking_What_It_Is_and_Why_It_Counts/links/5849b49608aed5252bcbe531/Critical-Thinking-What-It-Is-and-Why-It-Counts.pdf.

Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education*, *33*(8), 345-350. https://doi.org/10.3928/0148-4834-19941001-05

Facione, P. A., Sanchez, C. A., Facione, N. C., & Gainen, J. (1995). The disposition toward critical thinking. *The Journal of General Education*, *44*(1), 1-25. Retrieved June 15, 2020 from https://www.jstor.org/stable/27797240

Farquhar, J. D. (2012). *Case study research for business*. London: Sage Publications Ltd.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, *5*(1), 80-92. https://doi.org/10.1177%2F160940690600500107

Fisher, A. (2001). *Critical Thinking: An Introduction*. Cambridge: Cambridge University Press.

Fisher, A., & Scriven, M. (1997). *Critical thinking: Its definition and assessment*. Point Reyes, CA: Edgepress.

Flyvberg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, *12*(2), 219-245. https://doi.org/10.1177%2F1077800405284363

Flyvberg, B. (2011) Case Study. In: N. K. Denzin & Y. S. Lincoln (Eds.). *The Sage handbook of qualitative research*, 4. (pp. 301-316). London: SAGE Publications Ltd.

Fox, H. (1994). *Listening to the world: Cultural issues in academic writing*. Urbana, IL: National Council of Teachers of English. Retrieved December 23, 2016, from https://files.eric.ed.gov/fulltext/ED373331.pdf

Frick, B.L., Albertyn, R.M., & Bitzer, E.M. (2014). Conceptualising risk in doctoral education: navigating boundary tensions. In E.M. Bitzer, R.M. Albertyn, B.L. Frick, B. Grant & F. Kelly (Eds.), *Candidates, supervisors and institutions: Pushing postgraduate boundaries*. Stellenbosch: SunMedia. Retrieved July 18, 2019, from http://scholar.sun.ac.za/handle/10019.1/101646

Garside, C. (1996). Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills. *Communication Education, 45*(3), 212-227. https://doi.org/10.1080/03634529609379050

Gaunt, D. (1999). The practitioner as external examiner. *Quality in higher education*, *5*(1), 81-90. https://doi.org/10.1080/1353832990050108

Gee, J. P. (2015). *Social linguistics and literacies: Ideology in discourses* (5th ed.). Oxon: Routledge.

George-Williams, S., Carroll, M. R., Ziebell, A., Thompson, C., & Overton, T. (2019). Curtailing marking variation and enhancing feedback in large scale undergraduate chemistry courses through reducing academic judgement: a case study. *Assessment & Evaluation in Higher Education*, *44*(6), 881-893. https://doi.org/10.1080/02602938.2018.1545897

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: RoutledgeFalmer.

Glaser, E. M. (1941). *An experiment in the development of critical thinking*. New York: AMS.

Glaser, R., & Klaus, D. J. (1962). Proficiency measurement: Assessing human performance. In: R. M. Gagne (Ed.), *Psychological principles in system development*. (pp. 419-474). New York: Holt, Rinehart & Winston.

Golding, C. (2011). Educating for critical thinking: thought-encouraging questions in a community of inquiry. Higher Education Research & Development, 30(3), 357-370. http://dx.doi.org/10.1080/07294360.2010.499144

Golding, C. (2019). Discerning student thinking: a practical theoretical framework for recognising or informally assessing different ways of thinking. *Teaching in Higher Education, 24*(4), 478-492. https://doi.org/10.1080/13562517.2018.1491024

Grainger, P., Adie, L., & Weir, K. (2016). Quality assurance of assessment and moderation discourses involving sessional staff. *Assessment & Evaluation in Higher Education, 41*(4), 548-559. https://doi.org/10.1080/02602938.2015.1030333

Grainger, P., Purnell, K. & Zipf, R. (2008). Judging quality through substantive conversations between markers. *Assessment & Evaluation in Higher Education, 33*(2), 133-142. http://dx.doi.org/10.1080/02602930601125681

Gray, D. E. (2013). *Doing research in the real world*. London: Sage Publications Ltd.

Greenhalgh, T., & Taylor, R. (1997). How to read a paper: papers that go beyond numbers (qualitative research). *BMJ*, *315*(7110), 740-743. https://doi.org/10.1136/bmj.315.7110.740

Griffith, P. L., & Ruan, J. (2005). What Is Metacognition and What Should Be Its Role in Literacy Instruction? In S. E. Israel, C. C. Block, K. L. Bauserman, & K. Kinnucan-Welsch (Eds.), *Metacognition in Literacy Learning: Theory, Assessment, Instruction, and Professional Development*. (pp. 3-18). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Gu, Q. & Schwiesfurth, M. (2011). Rethinking University Internationalisation: Towards Transformative Change. *Teachers and Teaching, 17*(6), 611-617. https://doi.org/10.1080/13540602.2011.625110

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105-117.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic Controversies, Contradictions and Emerging Confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Sage Handbook of Qualitative Research*. (pp.191-216). Thousand Oaks, CA: Sage Publications, Inc.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is Originality in the Humanities and the Social Sciences?. *American Sociological Review*, *69*(2), 190-212. https://doi.org/10.1177/000312240406900203

Guo, F. & Shi, J. (2016). The relationship between classroom assessment and undergraduates' learning within Chinese higher education system. Studies in Higher Education, 41(4), 642-663. http://dx.doi.org/10.1080/03075079.2014.942274

Halonen, J. S. (1995). Demystifying Critical Thinking. *Teaching of Psychology, 22*(1), 75-81. https://doi.org/10.1207/s15328023top2201_23

Halonen, J. S., Smith, R. A., & Dunn, D. (2008). *Teaching critical thinking in psychology*. United Kingdom: Wiley-Blackwell.

Halpern, D. F. (1996). *Thinking critically about critical thinking*. London: Routledge.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455. https://doi.org/10.1037/0003-066X.53.4.449

Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed.). East Sussex: Psychology Press.

Halpern, D. F. & Sternberg, R. J. (2020). An Introduction to Critical Thinking: Maybe It Will Change Your Life. In: Sternberg, R. J. & Halpern, D. F. (Eds.), *Critical Thinking in Psychology* (2nd ed). (pp. 1-9). Cambridge: Cambridge University Press.

Hamilton, L., & Corbett-Whittier, C. (2013). *Using case study in education research*. London: Sage Publications Ltd.

Hammer, S. (2017). An examination of the instruction provided in Australian essay guides for students' development of a critical viewpoint. *Assessment & Evaluation in Higher Education, 42*(7), 1069-1081. https://doi.org/10.1080/02602938.2016.1224998

Hammer, S., & Griffiths, P. (2015). "Proposition Testing: A Strategy to Develop Critical Thinking for Essay Writing." In: M. Davies & R. Barnett (Eds.), *Palgrave Handbook of Critical Thinking in Higher Education*. (pp. 247–263). Basingstoke: Palgrave MacMillan. https://doi.org/10.1057/9781137378057_16

Hammer, S., Abawi, L., Gibbings, P., Jones, H., Redmond, P., & Shams, S. (2018). Developing a generic review framework to assure capstone quality. *Higher Education Research & Development, 37*(4), 730-743. https://doi.org/10.1080/07294360.2018.1453787

Hancock, D. R., & Algozzine, B. (2017). *Doing case study research: A practical guide for beginning researchers* (3rd ed.). New York: Teachers College Press.

Hannan, A., & Silver, H. (2006). On being an external examiner. *Studies in Higher Education*, *31*(1), 57-69. https://doi.org/10.1080/03075070500392300

Hackley, C. E. (1998). Social constructionism and research in marketing and advertising. *Qualitative Market Research: An International Journal*. https://doi.org/10.1108/13522759810235188/full/html

Hand, L., & Clewes, D. (2000). Marking the Difference: An investigation of the criteria used for assessing undergraduate dissertations in a business school. *Assessment & Evaluation in Higher Education, 25*(1), 5-21. http://dx.doi.org/10.1080/713611416

Handley, K., den Outer, B., & Price, M. (2013). Learning to mark: Exemplars, dialogue and participation in assessment communities. *Higher Education Research & Development*, *32*(6), 888-900. https://doi.org/10.1080/07294360.2013.806438

Harrell, M. (2011). Argument diagramming and critical thinking in introductory philosophy. *Higher Education Research & Development, 30*(3), 371-385. http://dx.doi.org/10.1080/07294360.2010.502559

Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual studies*, *17*(1), 13-26. Retrieved April 8, 2021, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.1304&rep=rep1&type=pdf

Haywood, P. & Wragg, E. (1982) Evaluating the literature, Rediguide 2. Nottingham: School of Education, University of Nottingham.

Healey, M. (2013). Assessing Capstone Projects and Dissertations. The University of Hong Kong Centre for the Enhancement of Teaching and Learning Wise Assessment Briefing No.4. Retrieved May 4, 2021, from https://www.cetl.hku.hk/teaching-learning-cop/wp-content/uploads/2015/08/wise-assessment-briefing4.pdf

Heijltjes, A., Van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning and Instruction*, *29*, 31-42. https://doi.org/10.1016/j.learninstruc.2013.07.003

Hennessey, M. G. (1999). Probing the dimensions of metacognition: Implications for conceptual change teaching-learning. *Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.* Retrieved June 15, 2020, from https://files.eric.ed.gov/fulltext/ED446921.pdf

Heywood, J. (2000). *Assessment in higher education: Student learning, teaching, programmes and institutions* (Vol. 56). London: Jessica Kingsley Publishers.

Hilberg, N. (2010). Is Originality an Appropriate Requirement for Undergraduate Publication?. *Honors in Practice, 6*, 57-60. Retrieved August 13, 2020, from https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1118&context=nchchip

Hinchliffe, G. W., & Jolly, A. (2011). Graduate identity and employability. *British Educational Research Journal*, *37*(4), 563-584. http://dx.doi.org/10.1080/01411926.2010.482200

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. *Educational psychology review*, *13*(4), 353-383. Retrieved April 10, 2018 from https://www.itma.vt.edu/courses/tel/resources/hofer(2001)_personal_epistemology_unit1.pdf

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational psychologist*, *39*(1), 43-55. https://doi.org/10.1207/s15326985ep3901_5

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of educational research*, *67*(1), 88-140. https://doi.org/10.2307/1170620

Holbrook, A., Bourke, S., Fairbairn, H., & Lovat, T. (2007). Examiner comment on the literature review in Ph. D. theses. *Studies in Higher Education*, *32*(3), 337-356. https://doi.org/10.1080/03075070701346899

Hornby, W. I. N. (2003). Assessing using grade-related criteria: a single currency for universities?. *Assessment & Evaluation in Higher Education*, *28*(4), 435-454. https://doi.org/10.1080/0260293032000066254

Hounsell, D. (2005). Contrasting conceptions of essay writing. In: F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (3rd Internet ed.). (pp. 106–125). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.

Howitt, D., & Cramer, D. (2014). *Introduction to Research methods in psychology* (4th ed.). Harlow: Pearson.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, *60*(2), 237-263. https://doi.org/10.2307/1170611

Hunter, K., & Docherty, P. (2011). Reducing variation in the assessment of student writing. *Assessment & Evaluation in Higher Education*, *36*(1), 109-124. https://doi.org/10.1080/02602930903215842

Ikuenobe, P. (2001). Teaching and Assessing Critical Thinking Abilities as Outcomes in an Informal Logic Course. *Teaching in Higher Education, 6*(1), 19-32. http://dx.doi.org/10.1080/13562510020029572

Jackson, N., & Shaw, M. (2006). Developing Subject Perspectives on Creativity in Higher education. In: N. Jackson, M. Oliver, M. Shaw & J. Wisdom (Eds.), *Developing Creativity in Higher Education: An Imaginative Curriculum*. (pp. 89–108). London: Routledge. Retrieved October 18, 2018, from http://www.creativeacademic.uk/uploads/1/3/5/4/13542890/__developing_subject_perspectives_on_creativity_in_higher_education.pdf

James, N., Hughes, C., & Cappa, C. (2010) 'Conceptualising, developing and assessing critical thinking in law'. *Teaching in Higher Education, 15*(3), 285-297. https://doi.org/10.1080/13562511003740858

James, R., McInnis, C., & Devlin, M. (2002). *Assessing learning in Australian universities: ideas, strategies and resources for quality in student assessment*. Melbourne, Vic: Centre for the Study of Higher Education.

Jansson, M., Carle, J., Gunnarsson, A., & Ekbrand, H. (2019). How experience affects assessment–a longitudinal study of assessment of university students' essays. *Studies in Higher Education*, *44*(4), 719-732. https://doi.org/10.1080/03075079.2017.1398227

Jawitz, J. (2009). Learning in the academic workplace: The harmonization of the collective and the individual habitus. *Studies in Higher Education*, *34*(6), 601-614. https://doi.org/10.1080/03075070802556149

Joffe, H. (2012). Thematic analysis. In: D. Harper & A. Thompson (Eds.), *Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners*. (pp. 209-223). Chichester: Wiley-Blackwell.

Johnson, P., & Duberley, J. (2000). *Understanding management research*. London: SAGE Publications Ltd.

Johnston, B. (2004). Summative assessment of portfolios: an examination of different approaches to agreement over outcomes. *Studies in Higher Education*, *29*(3), 395-412. https://doi.org/10.1080/0307507041000168264 6

Jonassen, D. H., & B. Kim, B. (2010). Arguing to Learn and Learning to Argue: Design Justifications and Guidelines. *Educational Technology Research and Development* 58(4): 439–457. https://doi.org/10.1007/s11423-009-9143-8

Jones, A. (2004). Teaching critical thinking: An investigation of a task in introductory macroeconomics. *Higher Education Research & Development*, *23*(2), 167-181. https://doi.org/10.1080/0729436042000206645

Jones, A. (2009). Redisciplining generic attributes: the disciplinary context in focus. *Studies in Higher Education, 34*(1), 85-100. https://doi.org/10.1080/03075070802602018

Jones, D. (2020). Conceptual Analysis of Critical Thinking in the 1990s. *International Journal of Management (IJM)*, *11*(3), 278-286. Retrieved June 15, 2020 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3573552

Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education, 39*(7), 840-852. http://dx.doi.org/10.1080/02602938.2013.875117

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, *2*(2), 130-144. https://doi.org/10.1016/j.edurev.2007.05.002

Kek, M. Y. C. A., & Huijser, H. (2011) The power of problem-based learning in developing critical thinking skills: preparing students for tomorrow's digital futures in today's classrooms. *Higher Education Research & Development, 30*(3), 329-341. http://dx.doi.org/10.1080/07294360.2010.501074

Kember, D. (2001). Beliefs about knowledge and the process of teaching and learning as a factor in adjusting to study in higher education. *Studies in Higher Education*, *26*(2), 205-221. https://doi.org/10.1080/03075070120052116

King, N. (2004). Using templates in the thematic analysis of text. In: C. Cassell & G. Symons (Eds.), *Essential Guide to Qualitative Methods in Organizational Research*. (pp. 256-270). London: SAGE Publications Ltd.

King, P. M., & Kitchener, K. S. (1994). *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults.* San Francisco: Jossey-Bass.

King, P. M., & Kitchener, K. S. (2002). The reflective judgment model: Twenty years of research on epistemic cognition. In: B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*. (pp. 37-62). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational psychologist*, *39*(1), 5-18. https://doi.org/10.1207/s15326985ep3901_2

King, N., Horrocks, C., & Brooks, J. (2019). *Interviews in qualitative research* (2nd ed.). London: SAGE Publications Limited.

Kitchener, K. S., & King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. *Journal of applied developmental psychology*, *2*(2), 89-116. https://doi.org/10.1016/0193-3973(81)90032-0

Kite, J., & Phongsavan, P. (2017). Evaluating standards-based assessment rubrics in a postgraduate public health subject. *Assessment & Evaluation in Higher Education*, *42*(6), 837-849. https://doi.org/10.1080/02602938.2016.1199773

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation review*, *31*(5), 415-439. https://doi.org/10.1177/0193841X07303318

Klenowski, V., & Wyatt-Smith, C. (2013). *Assessment for education: Standards, judgement and moderation*. London: Sage.

Knight, P. (2006). The local practices of assessment. *Assessment & Evaluation in Higher Education, 31*(4), 435-452. http://dx.doi.org/10.1080/02602930600679126

Knight, P., & Yorke, M. (2003). *Assessment, learning and employability*. Maidenhead: McGraw-Hill Education (UK).

Koole, M. (2012). A Social Constructionist Approach to Phenomenographic Analysis of Identity Positioning in Networked Learning. In: V. Hodgson, C. Jones, M. de Laat, D. McConnell, T. Ryberg & P. Sloep (Eds.). Proceedings of the 8th International Conference on Networked Learning 2012. (pp.1-8). Retrieved March 6, 2017, from https://auspace.athabascau.ca/handle/2149/3244

Kuhn, D. (1991). *The skills of argument.* Cambridge: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*(2), 155. https://doi.org/10.17763/haer.62.2.9r424r0113t670l1

Kuhn, D. (1999). A Developmental Model of Critical Thinking. *Educational Researcher, 28*(2), 16-25, 46. https://doi.org/10.2307/1177186

Kuhn, D., & Dean, Jr, D. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into practice*, *43*(4), 268-273. https://doi.org/10.1207/s15430421tip4304_4

Kurfiss, J. G. (1988). *Critical Thinking: Theory, Research, Practice, and Possibilities. ASHE-ERIC Higher Education Report No. 2, 1988*. ASHE-ERIC Higher Education Reports, The George Washington University, One Dupont Circle, Suite 630, Dept. RC, Washington, DC 20036-1183. Retrieved December 22, 2020, from https://eric.ed.gov/?id=ED304041

Kuzich, S., Groves, R., O'Hare, S., & Pelliccione, L. (2010). Building Team Capacity: Sustaining Quality in Assessment and Moderation Practices in a Fully Online Unit. *Paper presented at the ATN Assessment Conference 2010 University of Technology Sydney*. Retrieved December 23, 2019 from https://www.uts.edu.au/sites/default/files/Kuzich.pdf.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage Publications, Inc.

Kvale, S. (2007). Doing interviews. London: Sage Publications Ltd.

Lai, K. (2012). Assessing participation skills: online discussions with peers. Assessment & Evaluation in Higher Education, 37(8), 933-947. http://dx.doi.org/10.1080/02602938.2011.590878

Laming, D. (1990). The reliability of a certain university examination compared with the precision of absolute judgements. *The Quarterly Journal of Experimental Psychology*, *42*(2), 239-254. https://doi.org/10.1080%2F14640749008401220

Lamont, M. (2009). *How professors think: inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.

Lee, N., & Loton, D. (2015). Integrating Research and Professional Learning – Australian Capstones. *Council on Undergraduate Research Quarterly, 35*(4), 28-34. https://link.gale.com/apps/doc/A465166037/AONE?u=duruni&sid=AONE&xid=49e1101d

Lee, N., & Loton, D. (2019). Capstone purposes across disciplines. *Studies in Higher Education, 44*(1), 134-150. https://doi.org/10.1080/03075079.2017.1347155

Lee, B., Collier, P. M., & Cullen, J. (2007). Reflections on the use of case studies in the accounting, management and organizational disciplines. *Qualitative Research in Organizations and Management: An International Journal, 2*(3), 169-178. https://doi.org/10.1108/17465640710835337

Lee, H-J., Lee, J., Makara, K. A., Fishman, B. J., & Hong, Y-I. (2015). Does higher education foster critical and creative learners? An exploration of two universities in South Korea and the USA. *Higher Education Research & Development, 34*(1), 131-146. http://dx.doi.org/10.1080/07294360.2014.892477

Leong, P. A. (2013). Thinking critically: a look at students' critiques of a research article. *Higher Education Research & Development, 32*(4), 575-589. http://dx.doi.org/10.1080/07294360.2012.708322

Lincoln Y.S. & Guba E. (1985). *Naturalistic Inquiry*. Thousand Oaks, CA: Sage Publications, Inc.

Lincoln, D., & Kearney, M-L. (2019) Promoting critical thinking in higher education. *Studies in Higher Education, 44*(5), 799-800. https://doi.org/10.1080/03075079.2019.1586322

Lindgren, M., & Packendorff, J. (2009). Social constructionism and entrepreneurship. *International Journal of Entrepreneurial Behavior & Research, 15*(1), 25-47. https://doi-org.ezphost.dur.ac.uk/10.1108/13552550910934440

Lipman, M. (1988). Critical Thinking – What Can It Be?. *Educational Leadership, 46*(1), 38-43. Retrieved June 15, 2020, from https://journal.viterbo.edu/index.php/at/article/view/403

Lipman, M. (2003). *Thinking in education*. Cambridge: Cambridge University Press.

Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing critical thinking in higher education: the HEIghten™ approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education, 41*(5), 677-694. https://doi.org/10.1080/02602938.2016.1168358

Livingston, J. A. (1997). Metacognition: An Overview, 1-7. Retrieved July 6, 2020, from https://files.eric.ed.gov/fulltext/ED474273.pdf

Lok, B., McNaught, C., & Young, K. (2016). Criterion-referenced and norm-referenced assessments: compatibility and complementarity. *Assessment & Evaluation in Higher Education, 41*(3), 450-465. http://dx.doi.org/10.1080/02602938.2015.1022136

Lovitts, B. E. (2005). Being a good course-taker is not enough: a theoretical perspective on the transition to independent research. *Studies in higher education*, *30*(2), 137-154. https://doi.org/10.1080/03075070500043093

Lovitts, B. E. (2007). *Making the implicit explicit: Creating performance expectations for the dissertation*. Sterling, Virginia: Stylus Publishing, LLC.

Lovitts, B. E. (2008). The transition to independent research: Who makes it, who doesn't, and why. *The journal of higher education*, *79*(3), 296-325. https://doi.org/10.1080/00221546.2008.11772100

Luk, J., & Lin, A. (2015) Voices Without Words: Doing Critical Literate Talk in English as a Second Language. *TESOL Quarterly 49*(1), 67–91. https://doi.org/10.1002/tesq.161

Lundquist, R. (1999). Critical Thinking and the Art of Making Good Mistakes. *Teaching in Higher Education, 4*(4), 523-530. http://dx.doi.org/10.1080/1356251990040408

Luque, L. B. (2011). *Giving Reasons: A Linguistic-pragmatic Approach to Argumentation Theory*. Dordrecht: Springer.

Lumley, T. (2005). *Assessing second language writing: The rater's perspective*. Frankfurt: P. Lang. Retrieved January 11, 2016, from https://www.researchgate.net/profile/Tom_Lumley/publication/326107094_Assessing_Second_Language_Writing_The_Rater's_Perspective/links/5b395f620f7e9b0df5e45f22/Assessing-Second-Language-Writing-The-Raters-Perspective.pdf

MacLellan, E. (2004). How reflective is the academic essay?. *Studies in Higher Education*, *29*(1), 75-89. https://doi.org/10.1080/1234567032000164886

Macpherson, K., & Owen, C. (2010) Assessment of critical thinking ability in medical students. *Assessment & Evaluation in Higher Education, 35*(1), 41-54. http://dx.doi.org/10.1080/02602930802475471

Madill, A., Jordan, A., & Shirley, C. (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *British journal of psychology*, *91*(1), 1-20. https://doi.org/10.1348/000712600161646

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, *9*(3), 1-14. http://ojs.aishe.org/index.php/aishe-j/article/view/335

Mahdavi, M. (2014). An Overview: Metacognition in Education. *International journal of multidisciplinary and current research, 2(6)*, 529-535. Retrieved July 6, 2020, from http://ijmcr.com/wp-content/uploads/2014/05/Paper5529-535.pdf

Malsch, B., & Salterio, S. E. (2016). "Doing good field research": Assessing the quality of audit field research. *Auditing: A Journal of Practice & Theory*, *35*(1), 1-22. https://doi.org/10.2308/ajpt-51170

Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. *The lancet*, *358*(9280), 483-488. https://doi.org/10.1016/S0140-6736(01)05627-6

Marshall, C., & Rossman, G. B. (2016). *Designing Qualitative Research* (6th ed.). Thousand Oaks, Sage Publications, Inc.

Martinez, M. E. (2006). What is metacognition?. *Phi Delta Kappan*, *87*(9), 696-699. Retrieved June 15, 2020, from https://doi.org/10.1177/003172170608700916

Matthews, B., & Ross, L. (2010). Research Methods: A Practical Guide for the Social Sciences. London: Longman.

Maxwell, S. (2010). *Using Rubrics to Support Graded Assessment in a Competency Based Environment. Occasional Paper,* 1-25. Adelaide, Australia: National Centre for Vocational Education Research Ltd. Retrieved March 29, 2016, from https://files.eric.ed.gov/fulltext/ED509189.pdf

Mayer, R., & Goodchild, F. (1995). *The critical thinker*. Guilford: *Brown & Benchmark*.

McQuade, R., Kometa, S., Brown, J., Bevitt, D., & Hall, J. (2020). Research project assessments and supervisor marking: maintaining academic rigour through robust reconciliation processes. *Assessment & Evaluation in Higher Education*, *45*(8), 1181-1191. https://doi.org/10.1080/02602938.2020.1726284

McPeck, J. E. (1981). *Critical Thinking and Education.* Oxford: Martin Robertson.

Meadows, M., & Billington, L. (2005). A review of the literature on marking reliability. *London: National Assessment Agency*. Retrieved March 29, 2016, from file:///C:/Downloads/QCDA104983_review_of_the_literature_on_marking_reliability.pdf

Medland, E. (2016). Assessment in higher education: drivers, barriers and directions for change in the UK. *Assessment & Evaluation in Higher Education, 41*(1), 81-96. http://dx.doi.org/10.1080/02602938.2014.982072

Medland, E. (2019). 'I'm an assessment illiterate': towards a shared discourse of assessment literacy for external examiners. *Assessment & Evaluation in Higher Education*, *44*(4), 565-580. https://doi.org/10.1080/02602938.2018.1523363

Menéndez-Varela, J-L., & Gregori-Giralt, E. (2016). The contribution of rubrics to the validity of performance assessment: a study of the conservation–restoration and design undergraduate degrees. *Assessment & Evaluation in Higher Education, 41*(2), 228-244. http://dx.doi.org/10.1080/02602938.2014.998169

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education. Revised and Expanded from "Case Study Research in Education."*. San Francisco: Jossey-Bass.

Meyers, C. (1986). *Teaching Students to Think Critically. A Guide for Faculty in All Disciplines.* San Francisco: Jossey-Bass.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81-97. Retrieved March 29, 2016, from https://pure.mpg.de/rest/items/item_2364276_4/component/file_2364275/content

Mingers, J. (2000). What is it to be critical? Teaching a critical approach to management undergraduates. *Management Learning*, *31*(2), 219-237. https://doi.org/10.1177%2F1350507600312005

Mitchell, S. (1994). The teaching and learning of argument in sixth forms and higher education: Final report. *Hull: University of Hull, Centre for Studies in Rhetoric*.

Mitchell, S., & Riddle, M. (2000). Improving the quality of argument in higher education: Final report. *School of Lifelong Learning and Education, Middlesex University*.

Moon, J. (2008). *Critical Thinking An exploration of theory and practice*. Abingdon, Oxon: Routledge.

Moore, T. (2004). The critical thinking debate: how general are general thinking skills?. *Higher Education Research & Development*, *23*(1), 3-18. http://dx.doi.org/10.1080/0729360320001684 69.

Moore, T. (2011a). *Critical thinking and language: The challenge of generic skills and disciplinary discourses*. Bloomsbury Publishing. Accessed May 4, 2021, from https://books.google.co.uk/books?hl=en&lr=&id=6hMSBwAAQBAJ&oi=fnd&pg=PR3&dq=Critical+thinking+and+language:+The+challenges+of+generic+skills&ots=qrgCxfZlLy&sig=Yuwsk0WiONK8p06GtGAvjzzPCIU#v=onepage&q=Critical%20thinking%20and%20language%3A%20The%20challenges%20of%20generic%20skills&f=false

Moore, T. J. (2011b). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research & Development*, *30*(3), 261-274. https://doi.org/10.1080/07294360.2010.501328

Moore, T. (2013). Critical thinking: seven definitions in search of a concept. *Studies in Higher Education, 38*(4), 506-522. http://dx.doi.org/10.1080/03075079.2011.586995

Mulnix, J. W. (2012). Thinking critically about critical thinking. *Educational Philosophy and theory*, *44*(5), 464-479. https://doi.org/10.1111/j.1469-5812.2010.00673.x

Mulryan-Kyne, C. (2010). Teaching large classes at college and university level: Challenges and opportunities. *Teaching in Higher Education*, *15*(2), 175-185. https://doi.org/10.1080/13562511003620001

Namey, E., Guest, G., Thairu, L., & Johnson, L. (2008). Data reduction techniques for large qualitative data sets. In: G. Guest & K. M. MacQueen (Eds.), *Handbook for Team-Based Qualitative Research*. (pp. 137-161). Plymouth: Altamira Press.

Nelson, J. (2005). *Cultivating judgment: A sourcebook for teaching critical thinking*. Stillwater, OK: New Forums Press.

Neil, D. T., Wadley, D. A., & Phinn, S. R. (1999). A generic framework for criterion-referenced assessment of undergraduate essays. *Journal of Geography in Higher Education*, *23*(3), 303-325. https://doi.org/10.1080/03098269985263

Nicholas, M., & Raider-Roth, M. (2011). Approaches Used by Faculty to Assess Critical Thinking--Implications for General Education. *Online Submission, Paper presented at the ASHE Annual Conference, 2011, 1-30*. Retrieved June 7, 2016, from https://eric.ed.gov/?id=ED536592

Nightingale, P., Te Wiata, I., Toohey, S., Ryan, G., Hughes, C., & Magin, D. (1997). *Assessing learning in universities*. Kensington, NSW: University of New South Wales. Retrieved March 21, 2016, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.5749&rep=rep1&type=pdf

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods*, *16*(1), 1-13. https://doi.org/10.1177%2F1609406917733847

O'Connell, B., De Lange, P., Freeman, M., Hancock, P., Abraham, A., Howieson, B., & Watty, K. (2016). Does calibration reduce variability in the assessment of accounting learning outcomes?. *Assessment & Evaluation in Higher Education, 41*(3), 331-349. http://dx.doi.org/10.1080/02602938.2015.1008398

O'Donovan, B. (2017). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback. *Higher Education*, *74*(4), 617-633. https://doi.org/10.1007/s10734-016-0068-y

O'Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher education*, *9*(3), 325-335. https://doi.org/10.1080/1356251042000216642

O'Hagan, R., & Wigglesworth, G. (2015). Who's marking my essay? The assessment of non-native-speaker and native-speaker undergraduate essays in an Australian higher education context. *Studies in Higher Education, 40* (9), 1729-1747. https://doi.org/10.1080/03075079.2014.896890

Olsen, J., & Statham, A. (2005). Critical thinking in political science: Evidence from the introductory comparative politics course. *Journal of Political Science Education*, *1*(3), 323-344. https://doi.org/10.1080/15512160500261186

Olson, D. R., & Astington, J. W. (1993). Thinking about thinking: Learning how to take statements and hold beliefs. *Educational Psychologist*, *28*(1), 7-23. https://doi.org/10.1207/s15326985ep2801_2

Orr, S. (2007). Assessment moderation: constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education*, *32*(6), 645-656. https://doi.org/10.1080/02602930601117068

Orrell, J. 2008. Assessment beyond belief: The cognitive process of grading. In: A. Havnes & L. McDowell (Eds.), *Balancing dilemmas in assessment and learning in contemporary education*. (pp. 251–263). London: Routledge.

Ostorga, A. N. (2006). Developing teachers who are reflective practitioners: A complex process. *Issues in Teacher Education, 15*(2), 5-20. Retrieved April 10, 2018 from https://eric.ed.gov/?id=EJ796265

Packer, M. J. (2017). *The science of qualitative research* (2nd ed.). Cambridge: Cambridge University Press.

Pascarella, E.T., & Terenzini, P.T. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

PASSMORE, J. (1967). On teaching to be critical. In: R. S. Peters (Ed.), *The Concept of Education*. (pp.192-211). New York: Humanities Press.

Partington, J. (1994). Double-marking Students' Work. *Assessment & Evaluation in Higher Education, 19*(1), 57-60. https://doi.org/10.1080/0260293940190106

Pathirage, C., Haigh, R., Amaratunga, D., & Baldry, D. (2007). Enhancing the quality and consistency of undergraduate dissertation assessment. *Quality Assurance in Education, 15*(3), 271-286. https://doi-org/10.1108/09684880710773165

Paul, R. (1981). Teaching critical thinking in the" strong" sense: A focus on self-deception, world views, and a dialectical mode of analysis. *Informal Logic*, *4*(2), 2-7. https://doi.org/10.22329/il.v4i2.2766

Paul, R. (1990). *Critical Thinking Handbook: 4th-6th Grades. A Guide for Remodelling Lesson Plans in Language Arts, Social Studies, and Science*. Rohnert Park, CA: Center for Critical Thinking and Moral Critique, Sonoma State University.

Paul, R., & Elder, L. (2006). Critical thinking: The nature of critical and creative thought. *Journal of Developmental Education*, *30*(2), 34-35. Retrieved December 23, 2016, from https://search.proquest.com/docview/228409035/fulltextPDF/36EE660F713D49C5PQ/2?accountid=14533

Payan-Carreira, R., Cruz, G., Papathanasiou, I. V., Fradelos, E., & Jiang, L. (2019). The effectiveness of critical thinking instructional strategies in health professions education: a systematic review. *Studies in Higher Education, 44*(5), 829-843. https://doi.org/10.1080/03075079.2019.1586330

Peeters, M. J., Schmude, K. A., & Steinmiller, C. L. (2014). Inter-rater reliability and false confidence in precision: Using standard error of measurement within PharmD admissions essay rubric development. *Currents in Pharmacy Teaching and Learning*, *6*(2), 298-303. https://doi.org/10.1016/j.cptl.2013.11.014

Penkauskienė, D., Railienė, A., & Cruz, G. (2019) How is critical thinking valued by the labour market? Employer perspectives from different European countries. *Studies in Higher Education, 44*(5), 804-815. https://doi.org/10.1080/03075079.2019.1586323

Perkins, D., Jay, E., & Tishman, S. (1994). Assessing thinking: A framework for measuring critical-thinking and problem-solving. *US Department of Education Document no. NCES* 94-286, 65-112.

Perry, W. (1970*). Forms of intellectual and academic developments in the college years*. New York: Holt.

Phillips*, V., & Bond, C. (2004). Undergraduates' experiences of critical thinking. *Higher Education Research & Development*, *23*(3), 277-294. https://doi.org/10.1080/0729436042000235409

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: a review. *Educational Research, 42*(3), 237-249. https://doi.org/10.1080/001318800440579

Pope, R. (2005). *Creativity: Theory, history, practice*. London: Routledge.

Popham, W.J. (2005). *Classroom assessment: what teachers need to know* (4th ed.). New York: Pearson.

Price, M. (2005). Assessment standards: the role of communities of practice and the scholarship of assessment. *Assessment & Evaluation in Higher Education*, *30*(3), 215-230. https://doi.org/10.1080/02602930500063793

Price, M., & Rust, C. (1999). The experience of introducing a common criteria assessment grid across an academic department. *Quality in Higher Education*, *5*(2), 133-144. https://doi.org/10.1080/1353832990050204

Price, M., Carroll, J., O'Donovan, B., & Rust, C. (2011). If I was going there I wouldn't start from here: a critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education*, *36*(4), 479-492. https://doi.org/10.1080/02602930903512883

Price, M., O'Donovan, B., Rust, C., & Carroll, J. (2008). Assessment standards: a manifesto for change. *Brookes eJournal of Learning and Teaching*, *2*(3), 1-5. Retrieved October 5, 2018, from https://scholar.google.co.uk/scholar?hl=en&as_sdt=0%2C5&q=Assessment+Standards%3A+A+Manifesto+for+Change&btnG=

Prosser, M., & Webb, C. (1994). Relating the process of undergraduate essay writing to the finished product. *Studies in Higher Education*, *19*(2), 125-138. https://doi.org/10.1080/03075079412331381987

Puig, B., Blanco-Anaya, P., Bargiela, I. M., & Crujeiras-Pérez, B. (2019). A systematic review on critical thinking intervention studies in higher education across professional fields. Studies in Higher Education, 44(5), 860-869. https://doi.org/10.1080/03075079.2019.1586333

Punch, K.F. (2005). *Introduction to Social Research Quantitative and Qualitative Approaches* (2nd ed.). London: Sage Publications Ltd.

Quality Assurance Agency for Higher Education. (2014). The Frameworks for HE qualifications of UK degree-awarding bodies. Retrieved May 19, 2020, from https://www.qaa.ac.uk/quality-code/qualifications-and-credit-frameworks

Ramage, J. D., Callaway, M., Clary-Lemon, J., & Waggoner, Z. (2009). *Argument in composition*. Anderson: Parlor Press.

Rands, M., & Gansemer-Topf, A. M. (2016). Phenomenography: A methodological approach for assessment in student affairs. *Journal of Student Affairs Inquiry*, *1*(2), 1-22. Retrieved March 6, 2017, from https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1044&context=edu_pubs

Read, B., Francis, B., & Robson, J. (2005). Gender,'bias', assessment and feedback: Analyzing the written assessment of undergraduate history essays. *Assessment & Evaluation in Higher Education*, *30*(3), 241-260. https://doi.org/10.1080/02602930500063827

Rear, D. (2019) One size fits all? The limitations of standardised assessment in critical thinking. *Assessment & Evaluation in Higher Education, 44*(5), 664-675. https://doi.org/10.1080/02602938.2018.1526255

Reimann, N., Harman, K., Wilson, A., & McDowall, L. (2010). Learning to Assess in Higher Education: A Collaborative Exploration of the Interplay of Formal and Informal Learning in the Academic Workplace. *Paper presented at Higher Education Close Up Conference, Lancaster, UK, July 20–22*. Retrieved June 29, 2018, from http://nrl.northumbria.ac.uk/id/eprint/2872/1/Reimann%20N,%20Harman%20K,%20Wilson%20A,%20McDowell%20L%20-%20Learning%20to%20assess%20in%20higher%20education...%20conference%20paper.pdf

Remenyi, D. (2013, April). Case study research: The quick guide series. Reading: Academic Conferences and Publishing Limited.

Robson, C. (2002). *Real World Research*. Oxford: Blackwell.

Robson, C., & McCartan, K. (2016). *Real world research*. Chichester: John Wiley & Sons.

Rodriguez, C. M. (2009). The impact of academic self-concept, expectations and the choice of learning strategy on academic achievement: the case of business students, *Higher Education Research & Development, 28*(5), 523-539. http://dx.doi.org/10.1080/07294360903146841

Roohr, K., Olivera-Aguilar, M., Ling, G., & Rikoon, S. (2019) A multi-level modeling approach to investigating students' critical thinking at higher education institutions. *Assessment & Evaluation in Higher Education, 44*(6), 946-960. https://doi.org/10.1080/02602938.2018.1556776

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education*, *30*(3), 231-240. https://doi.org/10.1080/02602930500063819

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford review of education*, *13*(2), 191-209. https://doi.org/10.1080/0305498870130207

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144. https://doi.org/10.1007/BF00117714

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & evaluation in higher education*, *30*(2), 175-194. https://doi.org/10.1080/0260293042000264262

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education, 34*(2), 159-179. http://dx.doi.org/10.1080/02602930801956059

Sadler, D. R. (2010). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education*, *35*(6), 727-743. https://doi.org/10.1080/02602930902977756

Sadler, D. R. (2011). Academic freedom, achievement standards and professional identity. *Quality in Higher Education*, *17*(1), 85-100. https://doi.org/10.1080/13538322.2011.554639

Sadler, D. R. (2013). Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice, 20*(1), 5-19. http://dx.doi.org/10.1080/0969594X.2012.714742

Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, *67*(3), 273-288. https://doi.org/10.1007/s10734-013-9649-1

Saljo, R. (1997). Talk as data and practice—a critical look at phenomenographic inquiry and the appeal to experience. *Higher Education Research & Development*, *16*(2), 173-190. http://dx.doi.org/10.1080/0729436970160205

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: messages and meanings in the assessment of student learning. *Assessment & Evaluation in Higher Education*, *23*(4), 391-402. https://doi.org/10.1080/0260293980230406

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students* (6th ed.). Harlow: Pearson.

Schendel, R., & Tolmie, A. (2017) Beyond translation: adapting a performance-task-based assessment of critical thinking ability for use in Rwanda. *Assessment & Evaluation in Higher Education, 42*(5), 673-689. https://doi.org/10.1080/02602938.2016.1177484

Schommer, M. (1994). Synthesizing epistemological belief research: Tentative understandings and provocative confusions. *Educational psychology review*, *6*(4), 293-319. https://doi.org/10.1007/BF02213418

Schwandt, T. A. (2003). Three epistemological stances for qualitative inquiry: Interpretativism, hermeneutics and social constructionism. In: N. Denzin, & Y. Lincoln (Eds.), *The Landscape of Qualitative Research: Theories and issues*. (pp. 292-331). Thousand Oaks, CA: Sage.

Scott, M. (2000). Student, Critic and Literary Text: A discussion of 'critical thinking' in a student essay. *Teaching in Higher Education, 5*(3), 277-288. http://dx.doi.org/10.1080/713699140

Shafer, G. (2013). Higher level thinking, writing, and democracy among community college students. *Community College Journal of Research and Practice*, *37*(5), 382-387. https://doi.org/10.1080/10668921003609244

Shay, S. B. (2004). The assessment of complex performance: A socially situated interpretive act. *Harvard Educational Review*, *74*(3), 307-329. https://doi.org/10.17763/haer.74.3.wq16l67103324520

Shay, S. (2005). The assessment of complex tasks: a double reading. *Studies in Higher education*, *30*(6), 663-679. https://doi.org/10.1080/03075070500339988

Shay, S. (2008). Beyond social constructivist perspectives on assessment: the centring of knowledge. *Teaching in Higher education*, *13*(5), 595-605. https://doi.org/10.1080/13562510802334970

Siegel, H. (1985). Educating reason: Critical thinking, informal logic, and the philosophy of education. *Informal Logic*, *7*(2), 69-81. https://doi.org/10.22329/il.v7i2.2706

Silver, H., & Williams, S. (1996). Academic standards and the external examiner system. *Changing conceptions of academic standards*, 27-48.

Silver, H., Stennett, A., & Williams, R. (1995). *The external examiner system: Possible futures*. London: Quality Support Centre.

Silverman, D. (2015). *Interpreting qualitative data*. London: Sage Publications Ltd.

Simons, H. (1996). The paradox of case study. *Cambridge Journal of Education*, *26*(2), 225-240. https://doi.org/10.1080/0305764960260206

Simons, H. (2009). *Case study research in practice*. London: SAGE Publications Ltd.

Simpkins, W. S. (1987). The way examiners assess critical thinking in educational administration theses. Journal of Educational Administration, 25(2), 248–268. https://doi.org/10.1108/eb009935

Sin, S. (2010). Considerations of quality in phenomenographic research. *International Journal of Qualitative Methods*, *9*(4), 305-319. https://doi.org/10.1177/160940691000900401

Smith, C. (2012). Why should we bother with assessment moderation?. *Nurse Education Today, 32*(6), e45-e48. https://doi.org/10.1016/j.nedt.2011.10.010

Smith, E., & Coombe, K. (2006). Quality and qualms in the marking of university assignments by sessional staff: An exploratory study. *Higher Education*, *51*(1), 45-69. https://doi.org/10.1007/s10734-004-6376-7

Smith, D., Campbell, J., & Brooker, R. (1999). The impact of students' approaches to essay writing on the quality of their essays. *Assessment & Evaluation in Higher Education*, *24*(3), 327-338. https://doi.org/10.1080/0260293990240306

Spencer, L., Ritchie J., & O'Connor, W. (2003). Analysis: Practices, principles and processes. In: J. Ritchie & J. Lewis (Eds.). Qualitative Research Practice: A Guide for Social Sciences Students and Researchers. (pp. 199-218). London: Sage Publications Ltd.

Spicer, K. L., & Hanks, W. E. (1995). Multiple Measures of Critical Thinking Skills and Predisposition in Assessment of Critical Thinking. Paper presented at the Annual Meeting of the Speech Communication Association (81st, San Antonio, TX, November 18-21, 1995). Retrieved August 14, 2019, from https://files.eric.ed.gov/fulltext/ED391185.pdf

Stake, R. E. (2005). Qualitative Case Studies. In: N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (3rd ed.). (pp. 443-466). London: SAGE Publications Ltd.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342. http://dx.doi.org./10.1037/0022-0663.89.2.342

Starks, H., & Brown Trinidad, S. (2007). Choose your method: A comparison of phenomenology, discourse analysis, and grounded theory. *Qualitative health research*, *17*(10), 1372-1380. https://doi.org/10.1177/1049732307307031

Starman, A. B. (2013). The case study as a type of qualitative research. *Journal of Contemporary Educational Studies/Sodobna Pedagogika*, *64*(1), 28-43. Retrieved May 4, 2021 from https://d1wqtxts1xzle7.cloudfront.net/55511162/CaseStudy-with-cover-page.pdf?Expires=1620479568&Signature=BtLzOB~b6rr1OoYxpSgL8LCMthxtAsncmueIVMTb-WgcoVDarKxLbcupU58n7MugXTeMEVg8xpeU9N2CHpYaPjbkyT5eBGb9EVU1y7GyQ3UQqa~A551J7RoNhpGdtOp7Jy962jOV0o4tLhMpaBIRGaLnT1RN9VPkNdOxtGck7LyUFeW2jFL3wWOG4HNWh~ueyzsnsOnFdpuFtR9twyG4ROxK0xMlZd26Vrhz0RvSQCoA2GgfPaEpb6OgVcRR4XZPNIJ1Oiz8cBC2~AiamMtPy4hDDx5htZzbVaK6m~NfxbDLybykrG3fhd0riz8~sWaAb3~Qd-3nggc-kncZM7frHQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Stassen, M., Herrington, A., & Henderson, L. (2011). Defining critical thinking in higher education: Determining assessment fit. *To Improve the Academy, 30*, 126–141. https://doi.org/10.1002/j.2334-4822.2011.tb00653.x

Sternberg, R. J., & Kaufman, J. C. (2010). Constraints on Creativity: Obvious and Not So Obvious. In: J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge Handbook of Creativity*, 467–82. Cambridge: Cambridge University Press.

Sternberg, R. J., & Lubart, T. I. (1999). The Concept of Creativity: Prospects and Paradigms. In: R. J. Sternberg (Ed.), *Handbook of Creativity*. (pp. 3–15). Cambridge: Cambridge University Press.

Stupans, I., March, G., & Owen, S. M. (2013). Enhancing learning in clinical placements: reflective practice, self-assessment, rubrics and scaffolding. *Assessment & Evaluation in Higher Education*, *38*(5), 507-519. https://doi.org/10.1080/02602938.2012.658017

Swanwick, R., Kitchen, R., Jarvis, J., McCracken, W., O'Neil, R. & Powers, S. (2014) Following Alice: theories of critical thinking and reflective practice in action at postgraduate level. *Teaching in Higher Education, 19*(2), 156-169. http://dx.doi.org/10.1080/13562517.2013.836099

Tapper, J. (2004). Student perceptions of how critical thinking is embedded in a degree program. *Higher Education Research & Development, 23*(2), 199-222. http://dx.doi.org/10.1080/0729436042000206663

Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills'. *Research in Higher Education, 36*, pp.23–39. https://doi.org/10.1007/BF02207765

Thomas, G. (2010). Doing case study: Abduction not induction, phronesis not theory. *Qualitative inquiry*, *16*(7), 575-582. https://doi.org/10.1177/1077800410372601

Thomas, G. (2011a). A typology for the case study in social science following a review of definition, discourse, and structure. *Qualitative inquiry*, *17*(6), 511-521. https://doi.org/10.1177/1077800411409884

Thomas, G. (2011b). The case: generalisation, theory and phronesis in case study. *Oxford review of education*, *37*(1), 21-35. https://doi.org/10.1080/03054985.2010.521622

Thomas, G. (2013). From question to inquiry: operationalising the case study for research in teaching. *Journal of education for teaching*, *39*(5), 590-601. https://doi.org/10.1080/02607476.2013.852299

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, *15*(4), 663-682. https://doi.org/10.1007/s10763-016-9723-0

Tobin, G. A., & Begley, C. M. (2004). Methodological rigour within a qualitative framework. *Journal of advanced nursing*, *48*(4), 388-396. https://doi.org/10.1111/j.1365-2648.2004.03207.x

Todd, M., Bannister, P., & Clegg, S. (2004). Independent inquiry and the undergraduate dissertation: perceptions and experiences of final-year social science students. *Assessment & Evaluation in Higher Education, 29*(3), 335-355. https://doi.org/10.1080/0260293042000188285

Tsui, L. (2002). Fostering critical thinking through effective pedagogy. The Journal of Higher Education, 73(6), pp.740–763. https://doi.org/10.1080/00221546.2002.11777179

Tuckett, A. G. (2005). Applying thematic analysis theory to practice: A researcher's experience. *Contemporary nurse*, *19*(1-2), 75-87. https://doi.org/10.5172/conu.19.1-2.75

Tufford, L., & Newman, P. (2012). Bracketing in qualitative research. *Qualitative social work*, *11*(1), 80-96. https://doi.org/10.1177%2F1473325010368316

Tuononen, T., Parpala, A., & Lindblom-Ylänne, S. (2019). Graduates' evaluations of usefulness of university education, and early career success – a longitudinal study of the transition to working life. *Assessment & Evaluation in Higher Education, 44*(4), 581-595. https://doi.org/10.1080/02602938.2018.1524000

Universities UK. (2011). *Review of External examiner arrangements in universities and colleges in the UK: Final report and recommendations*. London: Universities UK.

van der Zanden, P. J. A. C., Denessen, E., Cillessen, A. H. N., & Meijer, P. C. (2018). Patterns of success: first-year student success in multiple domains. *Studies in Higher Education, 44*(11), 2081-2095. https://doi.org/10.1080/03075079.2018.1493097

van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching, 53*(1), 41–46. https://doi.org/10.3200/CTCH.53.1.41-48

Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice*, *6*(5), 6-7. http://dx.doi.org/10.5430/jnep.v6n5p100

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & health sciences*, *15*(3), 398-405. https://doi.org/10.1111/nhs.12048

Venning, J., & Buisman-Pijlman, F. (2013). Integrating assessment matrices in feedback loops to promote research skill development in postgraduate research projects. *Assessment & Evaluation in Higher Education*, *38*(5), 567-579. https://doi.org/10.1080/02602938.2012.661842

Verburgh, A. (2019). Effectiveness of approaches to stimulate critical thinking in social work curricula. *Studies in Higher Education, 44*(5), 880-891. https://doi.org/10.1080/03075079.2019.1586336

Walker, P., & Finney, N. (1999). Skill Development and Critical Thinking in Higher Education. *Teaching in Higher Education, 4*(4), 531-547. http://dx.doi.org/10.1080/1356251990040409

Walvoord, B. E., & Anderson, V. J. (2010). *Effective grading: A tool for learning and assessment in college* (2nd ed.). San Francisco: Jossey-Bass.

Warren, C. A., & Karner, T. X. (2015). *Discovering qualitative methods: Ethnography, interviews, documents, and images* (3rd ed.). Oxford: Oxford University Press.

Webster, F., Pepper, D., & Jenkins, A. (2000). Assessing the Undergraduate Dissertation. *Assessment & Evaluation in Higher Education, 25*(1), 71-80. http://dx.doi.org/10.1080/02602930050025042

Weigle, S. C. 2002. Assessing Writing. Cambridge: Cambridge University Press.

Wellington, J. (2013). Searching for 'doctorateness'. *Studies in Higher Education*, *38*(10), 1490-1503. https://doi.org/10.1080/03075079.2011.634901

Wendt, M., & Åse, C. (2015). Learning dilemmas in undergraduate student independent essays. *Studies in Higher Education, 40*(5), 838-851. http://dx.doi.org/10.1080/03075079.2013.842967

West, E. J. (2004). Perry's legacy: models of epistemological development. *Journal of adult development*, *11*(2), 61-70. https://doi.org/10.1023/B:JADE.0000024540.12150.69

Whalley, W. B. (2016). Evaluating student assessments: the use of optimal foraging theory. *Assessment & Evaluation in Higher Education, 41*(2), 183-198. http://dx.doi.org/10.1080/02602938.2014.991909

Wiliam, D. (1996). Standards in examinations: a matter of trust?. *The Curriculum Journal*, *7*(3), 293-306. https://doi.org/10.1080/0958517960070303

Williams, R. (1976) *Keywords*. Collins: London

Williams, L., & Kemp, S. (2019). Independent markers of master's theses show low levels of agreement. *Assessment & Evaluation in Higher Education*, *44*(5), 764-771. https://doi.org/10.1080/02602938.2018.1535052

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach?. *Arts Education Policy Review*, *109*(4), 21-32. https://doi.org/10.3200/AEPR.109.4.21-32

Wilson, A. N. & Howitt, S. M. (2018). Developing critical being in an undergraduate science course. *Studies in Higher Education, 43*(7), 1160-1171. https://doi.org/10.1080/03075079.2016.1232381

Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for academic purposes*, *11*(2), 145-154. https://doi.org/10.1016/j.jeap.2011.11.001

Winter, R., Griffiths, M., & Green, K. (2000). The'Academic' Qualities of Practice: What are the criteria for a practice-based PhD?. *Studies in higher education*, *25*(1), 25-37. https://doi.org/10.1080/030750700115993

Wolfe, E. W., Kao, C. W., & Ranney, M. (1998). Cognitive differences in proficient and nonproficient essay scorers. *Written Communication*, *15*(4), 465-492. https://doi.org/10.1177/0741088398015004002

Woolf, H. (2004). Assessment criteria: reflections on current practices. *Assessment & Evaluation in Higher Education*, *29*(4), 479-493. http://dx.doi.org/10.1080/02602930310001689046

Yates, C., Partridge, H., & Bruce, C. (2012). Exploring information experiences through phenomenography. *Library and Information Research*, *36*(112), 96-119. Retrieved March 6, 2017, from https://www.lirgjournal.org.uk/index.php/lir/article/view/496

Yates, S.J. (2004). *Doing Social Science Research*. London: Sage Publications Ltd.

Yin, R. K. (2011). *Applications of case study research* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Ylonen, A., Gillespie, H., & Green, A. (2018). Disciplinary differences and other variations in assessment cultures in higher education: exploring variability and inconsistencies in one university in England. *Assessment & Evaluation in Higher Education*, *43*(6), 1009-1017. https://doi.org/10.1080/02602938.2018.1425369

Yorke, M. (2008). *Grading student achievement in higher education: Signals and shortcomings*. Abingdon: Routledge.

Yorke, M. (2010). How finely grained does summative assessment need to be?. *Studies in Higher Education, 35*(6), 677-689. http://dx.doi.org/10.1080/03075070903243118

Yorke, M. (2011). Summative assessment: dealing with the 'measurement fallacy'. *Studies in Higher Education, 36*(3), 251-273. http://dx.doi.org/10.1080/03075070903545082

Yorke, M., Bridges, P., & Woolf, H. (2000). Mark distributions and marking practices in UK higher education: some challenging issues. *Active learning in higher education*, *1*(1), 7-27. https://doi.org/10.1177/1469787400001001002

Young, R. A., & Collin, A. (2004). Introduction: Constructivism and social constructionism in the career field. *Journal of vocational behavior*, *64*(3), 373-388. https://doi.org/10.1016/j.jvb.2003.12.005

Zucker, D. M. (2009). How to Do Case Study Research. Teaching Research Methods in the Social Sciences, 2. Retrieved August 6, 2019, from https://scholarworks.umass.edu/nursing_faculty_pubs/2