Learning about academic writing through holistic peer assessment

Natalie D. Usher

Department of Education & Wolfson College, University of Oxford

Supervisors: **Dr Therese N. Hopfenbeck** and **Dr Kathleen M. Quinlan**

Thesis submitted in part-fulfilment of the requirements

for the degree of Doctor of Philosophy in Education

April 2018

Acknowledgements

First of all, thank you to the students and tutors who made this study possible – for being so generous with your time and enthusiastic in your participation.

Thank you to the ESRC for the studentship that enabled me to conduct this research, and for funding an Overseas Institutional Visit to the University of Victoria, Canada. I'd also like to express my appreciation to OUCEA and Wolfson College for conference funding.

Thank you to my supervisors, Therese N. Hopfenbeck and Kathleen M. Quinlan for your support and encouragement. I've learned a lot from you both. I'm grateful you could continue to supervise me after your move to Kent, Kathleen.

Thank you to OUCEA colleagues and friends, past and present, and in particular Jo Hazell, Carol Brown, Yasmine El Masri and Diana Ng for your support and friendship.

Thank you to Dr Allyson Hadwin and her lab at UVic – and especially Sarah Davis, Aishah Bakhtiar, Becca Edwards and Mariel Miller. My visit to UVic was so inspiring and a real highlight of the DPhil.

Thank you to everyone who has helped me learn how to manage the elbow condition and/or provided support. In particular, I'd like to thank Jan Morris, Emma Deeny, Jo O'Donovan, Georgie Heywood, and Lauren McLeod. I was extremely fortunate to pick Wolfson as my college. Special thanks to Fiona Wilkes, librarian.

Thank you to Nadri Boswell – for your patient and kind practical support with coding, and for gardening chat and chocolate appreciation.

Thank you to my DPhil friends. In particular, Emma Abotsi, Sina Fackler and Christine Paget – I'm so proud that our MSc team is now about to complete the DPhil. And to the lovely CoNi – Jill Boggs, Moon Cho, and Susie Davis. Thank you for tea, wit, strength, and inspiration.

Thank you to my family for your support. A huge thank you to Sue and Steve Usher for kilos of chocolate and love, and for starting this whole thing by teaching me to read early (Mum) and sharing lots of good books (Dad). I'm really thankful for the support of my wider family, who happily grew in number during this degree – Andy, Britt, Trish, Tim, Rob, Chris, Rich, Irina, and the snowman. I am also extremely grateful for Gatita's writing companionship and helpful insistence on regular breaks.

Finally, thank you to my husband, Stu, who can attest that even when you research writing, it's *still* a highly complex and demanding process. Thank you for your unfailing patience, sense of humour and love.

Abstract

Learning about academic writing through holistic peer assessment

Natalie Usher, DPhil Education, Wolfson College, Hilary Term 2018

While there is a consensus among researchers that assessment should and can serve learning, there is less understanding of *how* it supports learning at a fine-grained level. This thesis uses design-based research to investigate the role of comment-only, holistic peer assessment in writing development. The theory of action synthesises Sadler's accounts of learning through assessment (1989, 2010) with Winne and Hadwin's (1998, 2008) model of self-regulated learning. It is theorised that participating in peer assessment helps students to develop evaluative expertise, which in turn enriches task perceptions, metacognitive standards and ultimately large-scale adaptation: the changes students employ in subsequent essays. Drawing on the theory of action, I designed a series of workshops for first-year English Literature students learning to write examination essays. The thesis reports on the first of two iterations. 21 participants assessed and discussed example essays; criteria were not pre-determined but emerged from discussion of four examples. Students then wrote a timed essay, assessed three peer pieces and received three reviews. A range of data was generated during the workshops, including written comments, reflections and questionnaires. Ten case study writers also took part in pre- and post-workshops writing tasks, think-aloud protocols and interviews.

To trace the development of students' evaluative expertise, I coded inductively students' talk and comment about writing. Visualising the connections between emergent codes reveals writing quality as a complex web of criteria, with the essay question at the centre. There was a strong overlap between the official Faculty assessment criteria and the codes emerging from student data. However, students also frequently commented on procedural aspects of writing such as introductions and conclusions, which are left tacit or latent in Faculty criteria. Post-workshops, students' own metacognitive standards became increasingly reader-oriented and question-focussed, and these procedural aspects of writing drove the adaptations they made to their approach. I use rich, in-depth case study data to trace how, why, and when students made such adaptations. I also examine the role of peer feedback, which rather than offering new information, often verified or complemented the judgements students formed of their own writing during the workshops. The thesis thus illuminates processes involved in learning through assessment. It also shows that peer assessment is a practicable way of developing within the discipline both evaluative expertise and writing, which are key to lifelong learning.

Table of contents

Abstract 3

List of appendices 8

List of figures, boxes and tables 9

Chapter 1: Introduction 11

1.1 The practical problem and the theoretical problem 11

1.2 Overview of thesis chapters 13

Chapter 2: Literature Review 26

2.1 Chapter introduction: the practical problem 26

2.2 Defining writing 31

2.3 Formative assessment in higher education 35

2.4 Formative assessment and writing 37

2.4.1 Writing quality 38

2.4.2 Measurement influences on formative assessment practices 40

2.5 Sadler's proposal 44

2.5.1 Comparing exemplar use and peer assessment 45

2.6 Peer assessment and writing 50

2.7 How might students learn from peer assessment of writing? 54

2.7.1 Taking a readerly perspective 54

2.7.2 Learning about genre 57

2.7.3 Observing others' performances 60

2.8 Receiving feedback 62

2.9 Transitioning to university writing in English Literature 67

2.10 Chapter conclusion 68

Chapter 3: Theory of action 70

3.1 Chapter introduction 70

3.2 Theory of action 71

3.3. Background to the theory of action 73

3.4 Research questions 80

3.5 Conditions 83

3.6 Self-efficacy beliefs 85

3.7 Task perceptions 87

3.8 Standards 92
3.9 Large-scale adaptation 98
3.10 Chapter conclusion 101

Chapter 4: Research Design and Methodology 102
4.1 Chapter introduction 102
4.2 Design-based research approach 104
4.3 Setting for the research 108
4.4 Sampling and participants 111
4.5 Designing the workshops 114
4.6 Procedure 117
4.7 Research methods: during the intervention 121
4.7.1 Self-efficacy questionnaire 123
4.7.2 Regulation of Learning Survey (RLQ) 123
4.8 Research methods: pre-and post-intervention 124
4.8.1 Think-aloud protocol and timed essay 125
4.8.2 Post-writing interviews 128
4.9 Reflexivity 129
4.10 Ethics 130
4.10.1 Procedural ethics: CUREC application 130
4.10.2 Ethical issues 131
4.10.3 Sharing findings with participants 135
4.11 Analytical strategy 136
4.11.1 Transcription and preparation 138
4.11.2 Analytical memos 140
4.11.3 Data-driven coding: criteria 140
4.11.4 Visualisations and SRL coding 144
4.11.5 Analytical strategies for each research question 149
4.11.6 Analysis of essays 154
4.12 Limitations of the research design and methods 156
4.13 Chapter conclusion 157

Chapter 5: What students notice about writing and choose to adapt **159**
5.1 Introduction 159
5.2 Identifying criteria in students' talk, comment and feedback on writing 161
5.3 How the coding relates to the English Faculty criteria 169
5.4 Student group criteria 173
5.5. The challenge of exam writing 179
5.6 Overview of large-scale adaptations 182
5.7 Barriers to adapting writing 190
5.8 What is enough adaptation? 191
5.9 Self-efficacy 196
5.10 Chapter conclusion 203

Chapter 6: When criteria become standards for writing **205**
6.1 Introduction 205
6.2 Definition and operationalisation of standards 206
6.3 Where do standards come from? 209
6.4 Which criteria are used as standards? 214
6.5 Do standards change over the course of the study? 219
6.6 Summary of coding findings 224
6.7 Focal case: Tessa 225
6.8 Tessa's planning at Interview 1 226
6.9 Workshops: Planning for structure 228
6.10 Tessa's planning at Interview 2: Structure as restrictive 230
6.11 Other standards for writing: Tessa's 'boxes' 235
6.12 Is it enough to tick 'boxes'? 237
6.13 Chapter conclusion 240

Chapter 7: Task perceptions **243**
7.1 Introduction 243
7.2 Developing task perceptions: an overview 247
7.3 Focal case: Ben 252
7.4 During workshops: 'DON'T IGNORE THE QUOTATION' 258
7.5 Post-workshops 260
7.6 Chloe's task perceptions pre-workshops 264
7.7 Chloe's task perceptions post-workshops 267
7.7 Chapter conclusion 270

Chapter 8: Feedback, or 'new ideas about things I was already aware of' **272**
8.1 Introduction 272
8.2 What do students comment on when giving feedback? 276
8.3 What types of comments do students make when giving feedback? 278
8.4 How do participants respond to feedback? 284
8.5 What defines helpful feedback? 286
8.7 What makes feedback unhelpful? 290
8.8 Agreeing with but not implementing feedback 294
8.9 Feedback as a 'boost' to self-efficacy 298
8.10 Chapter conclusion 300

Chapter 9: Discussion **303**
9.1 Introduction 303
9.2 Revisiting the theory of action 304
9.3 Evaluative expertise 309
9.4 Shifting standards 313
9.5 Task perceptions, task compliance and creativity 318
9.6 The theoretical problem revisited 323
9.7 Evaluative expertise and the role of feedback 327
9.8 Self-efficacy 331
9.9 The practical problem revisited 333
9.10 Conclusions 338

References **340**

Appendices **366**

Word count: 99,964

List of Appendices

A: Workshop materials	367
B: Questionnaires: demographics, formative assessment, workshop evaluation	387
C: Self-efficacy questionnaire and results	393
D: Adapted Regulation of Learning Questionnaire	400
E: Think-aloud, writing and interview schedule	404
F: CUREC application, letter to participants and consent form	408
G: Ethically important moments	425
H: Extract from memo on large-scale adaptation	430
I: Criteria coding guide	440
J: SRL coding guide	485
K: Layers of task perceptions	517
L: Feedback coding guides	520
M: Cross-tabulation showing intersections of criteria codes	529
N: Standards across phases of learning	533
O: Extract from Tessa's think-aloud protocol at Interview 1	534
P: Tessa's planning	535
Q: Tessa's introductions	538

List of figures, boxes and tables

Figures

Figure 1: Theory of action 72

Figure 2: Timeline and overview of data collection 103

Figure 3: Designing the workshops 114

Figure 4: Development of criteria coding guide 141

Figure 5: Example of visualisation process (Chloe) 146

Figure 6: Digitised visualisation (Chloe) 147

Figure 7: Development of SRL coding guide 148

Figure 8: Analytical process for Research Question 2 150

Figure 9: Analytical process for Research Question 4 152

Figure 10: Visualisation of criteria coding 165

Figure 11: Visualisation: subset of the most frequent criteria 166

Figure 12: Visualisation showing how Faculty criteria map onto coding 172

Figure 13: Tessa's annotations and additions on Example Essay 4, Workshop 2 233

Figure 14: Ben's question and introduction at pre-workshops 252

Figure 15: Ben's addition to the conclusion of Example Essay 1 259

Figure 16: Ben's question and introduction at post-workshops 261

Figure 17: Chloe's question and introduction at pre-workshops 266

Figure 18: Chloe's question and introduction at post-workshops 269

Figure 19: Feedback to Phoenix on his conclusion 295

Figure 20: Theory of action revisited 306

Boxes

Box 1: Research Questions 106

Box 2: Timeline for analysis 139

Tables

Table 1: Overview of research questions and methods 106

Table 2: Participants by college 112

Table 3: Participants by gender 113

Table 4: Breakdown of pre-university schooling – countries 113

Table 5: Breakdown of pre-university institution types 114

Table 6: Workshops procedure 118

Table 7: Overview of main qualitative dataset 137

Table 8: Overview of research questions and analytical methods 150

Table 9: Overview of criteria coding 163

Table 10: Mapping English Faculty criteria onto the coding guide 171

Table 11: Mapping workshop groups' criteria onto the English Faculty criteria and coding guide 175

Table 12: Overview of large-scale adaptations at Workshop (W) essay, and Interview 2 essay 184

Table 13: Participants for whom there were no adaptations linked to workshop learning 186

Table 14: Introduction word counts 188

Table 15: Self-efficacy: Wilcoxon signed-rank test results 198

Table 16: References to criteria codes 216

Table 17: Standards used at Time 1 and Time 2 221

Table 18: References to *clarity* in Tessa's data 226

Table 19: Functions and uses of feedback 274

Table 20: References to criteria codes in annotations and end comments on peer essays 277

Table 21: Types of comment in annotations and end comments on peer essays 280

Table 22: Cross-tabulation showing coding results for type of feedback comment by criterion 282

Table 23: Types of annotations made on feedback received 285

Table 24: Responses to praise, problem and solution comments 286

Table 25: Types of comment across subsets of feedback data 287

Chapter 1: Introduction

1.1 The practical problem and the theoretical problem

There is now a consensus that assessment should serve learning, and involve students not only as subjects but participants. However, there is less knowledge and consensus about which forms of assessment most appropriately support discipline and domain-specific learning processes, and this is particularly the case for writing. This study brings together multiple literatures to investigate the role of comment-only, holistic peer assessment in writing development, and in doing so addresses a practical issue and a theoretical issue. On a practical level, the study proposes that embedding peer assessment into disciplinary teaching may support writing development in universities where no formal writing instruction is offered. On a theoretical level, the study offers insights into how students learn from participating in formative assessment practices, contributing to a field where findings tend to focus on learning outcomes to the neglect of domain-specific learning processes.

To address these issues, the study evaluates a series of peer assessment writing workshops designed for first-year undergraduates in English Literature. The transition to university is an apt setting for the study as students are challenged to adapt their writing, but may not yet have the knowledge and strategies to do so. The workshops were designed using principles proposed by Sadler (2009a, 2010), as well as insights from the writing literatures. It is argued that Sadler's proposal for holistic peer assessment is particularly suited to support learning about writing complex, knowledge-rich, ill-defined tasks such as literary essays. Rather than applying pre-determined criteria, students read, appraise and discuss a range of example essays to develop their own evaluative expertise. In this study, they then handwrote a timed essay, before writing and receiving comment-only feedback on an anonymous basis. To bridge the fields of writing and formative assessment and offer insights into how students proactively manage such writing tasks,

the study uses a theory of action that draws on self-regulated learning (SRL) theory. This novel synthesis, which applies the Winne and Hadwin (1998, 2008) model of SRL to Sadler's (2009a, 2010) proposal, enables an examination on an intra-personal level of how peer assessment may enrich three aspects of self-regulation which contribute to learning and writing. These are: task perceptions, metacognitive standards and ultimately, large-scale adaptation, or intentional and strategic changes made from essay to essay. It also considers the impact on students' self-efficacy for writing.

The study was designed to tackle an over-arching research question: (how) does students' self-regulation of writing develop in the context of peer assessment writing workshops? The thesis specifically addresses five sub-questions:

> RQ1. During the workshops, which criteria do students use to evaluate written essay products?
>
> RQ2. Which criteria do students use as standards to self-regulate their essay writing processes?
>
> RQ3. (How) do students' task perceptions shift over the course of the study?
>
> RQ4. (How) does peer feedback contribute to students' self-regulation of writing?
>
> RQ5. What is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing?

What follows is an overview of the thesis chapters, beginning with the literature review. This chapter should indicate to the reader how the research questions above are addressed.

1.2 Overview of thesis chapters

Chapter 2 – the literature review – begins by discussing the practical problem addressed by the thesis. The transition to university is often highly challenging for two reasons: students are expected to write in new genres and synthesise more challenging reading, and there is less teacher regulation or direction of writing and learning. In contrast to other countries, universities in England usually operate on a study skills model: rather than offering academic writing instruction as part of teaching, support is only provided for certain groups, such as international students and those with disabilities (Lea and Street, 1998; Wingate, 2006). For most students, the main source of learning about writing is lecturer feedback on final pieces of work. However, feedback is often hard for students to interpret and use because the metalanguage used to refer to the often tacit and variable conventions of academic writing is slippery and opaque (Wingate, 2012). Assessment criteria pose similar problems (Sadler, 2014).

Feedback and assessment criteria become more helpful to students when used alongside other strategies which position students as active participants in assessment practices, as in Assessment for Learning, widely used in secondary and primary schools internationally (Black and Wiliam, 2009). This study draws on Sadler's (2010) proposal for formative assessment to better support learning at university by shifting in emphasis from feedback to training students to appraise work holistically. By developing evaluative expertise, students should be in a better position to self-monitor work as they produce it (Sadler, 1989). The value of Sadler's theorising about evaluative expertise is that it re-frames formative assessment as a process that is central to learning and teaching. This is in line with other theorisations of formative assessment such as Black and Wiliam (2009), Allal (2010) and Pryor and Crossouard (2008). However, in common with other theories of formative assessment, Sadler's proposal lacks insight into *how* students might self-monitor their

work. The literature also lacks attention to which types of formative assessment practice may most appropriately support domain-specific learning (Bennett, 2011).

The thesis uses two bodies of theory to consider how Sadler's proposal may work. First, the literature review chapter draws on the writing literatures to examine how assessment may support writing development. There is relatively little empirical research tracing the impact and processes involved in formative assessment and writing (Graham, Harris and Hebert, 2011). Peer assessment studies show promise, particularly that students learn about writing through giving feedback as well as receiving (Cho and Cho, 2011; Cho and MacArthur, 2011). However, again, this research lacks much detail on *how* students develop their writing through assessment. Examining practices reported in the literature, the influence of measurement rather than learning theory is apparent, and the impact on students is discussed and contextualised with the wider assessment literature. To unpick Sadler's proposal and review the empirical evidence for holistic peer assessment, the study examines related research on exemplars, and then looks at the writing theory more closely, in particular the role of reading like a reader in writing development, genre knowledge, and observing others' writing.

The second body of literature used to frame Sadler's proposal is self-regulated learning (SRL), which acts as a bridge between the writing and formative assessment literatures. Hacker, Keener and Kircher's (2009) definition of writing as applied metacognition is used to make the initial connection as it highlights the commonalities between the fields. SRL is then used to explain how peer assessment may help students develop their writing. Broadly, 'SRL is viewed as *proactive* processes that students use to acquire academic skill, such as setting goals, selecting and deploying strategies, and self-monitoring one's effectiveness, rather than as a reactive event that happens to students due to impersonal forces' (Zimmerman, 2008, pp.166–167). Theories of SRL

break learning down into phases, and offer accounts of metacognitive monitoring which resonate with Sadler's descriptions of learning through assessment (e.g Zimmerman, 2000; Winne and Hadwin, 1998). As well as subsuming metacognition, SRL also considers the interaction with the environment, such as the classroom, and is multi-dimensional, recognising that factors such as motivation, self-efficacy and affect are also self-regulated (Dinsmore et al., 2008; Hadwin, Järvelä, and Miller, 2011; Winne and Hadwin, 2008).

Chapter 3 presents the theory of action for the study, which uses Winne and Hadwin's (1998, 2008) model of self-regulated learning to explain how peer assessment as proposed by Sadler (2009a, 2010) may support writing. Building on the metacognitive definition of writing used in the literature review (Hacker et al. 2009), it also brings insights from Galbraith's (2009) dual-process theory of writing as involving both knowledge retrieval and knowledge constituting, or spontaneous discovery during the act of writing. Such processes can conflict, particularly under time pressure, and throw up self-regulatory challenges.

The theory of action proposes that by participating in peer assessment, writers develop evaluative expertise, or rich knowledge about quality and the salience of criteria which enables students to make qualitative judgements on work. Students develop their evaluative expertise in three concepts in particular: task compliance, or what it means to meet the requirements of the task, quality and criteria (Sadler, 2010). Seeing authentic responses to questions also informs students' task perceptions, which become more closely aligned with what teachers value. Evaluative expertise enriches the metacognitive standards students use to plan and monitor their writing. After a writing episode, students decide whether and how to adapt their writing in future essays, and peer assessment thus informs these decisions. So, while the study does investigate some aspects of within-task self-regulation, the main emphasis is on how participants' self-regulation of

writing develops from episode to episode. This level of granularity matches the focus in the formative assessment literature on learning over time. The theory of action therefore does not operate at the micro-trace level that much self-regulated learning research conducted online uses at present (e.g. Winne and Muis, 2011; Greene, Hutchison, Costa and Crompton, 2012).

Next, there is in-depth discussion of aspects of self-regulated learning important to the thesis: conditions, task perceptions, metacognitive standards and large-scale adaptation. In the first phase of SRL, the student defines (Winne and Hadwin, 1998) or perceives (Winne and Hadwin, 2008) the task. Task perceptions are shaped by conditions such as prior learning, and the task instructions themselves, and play an important role in self-regulating learning as the plans and strategies that follow depend on such task perceptions. The question/quotation that participants select is used – or not – to shape, or frame or even transform the student's literary knowledge. Task perceptions are important because they operate on explicit, implicit and socio-contextual levels, and the latter levels cannot be deduced from instructions alone (Hadwin and Winne, 2012). Developing evaluative expertise through activities such as peer assessment should influence students' task perceptions, enabling them to see aspects of the task beyond the explicit.

Metacognitive standards are the criteria used by students to monitor and control their learning or writing in the moment (Winne and Hadwin, 1998). Unlike broader educational or academic standards, which are set by communities of experts, published, and subjected to public scrutiny (Cresswell, 2000), metacognitive standards are personal, often tacit and idiosyncratic. Students' standards are 'multifaceted' and represent the 'optimal end state' of the current phase of learning, according to the learner's beliefs (Greene and Azevedo, 2007, p.336). While teachers may assume that where assessment criteria are provided, all are adopted and used to self-regulate learning, writers may not understand or agree with the criteria given, or know what a criterion

looks like in practice. In a subject area such as English, criteria are often fuzzy, overlapping and interlinked, making them difficult to define and agree on. By participating in peer assessment, students develop their evaluative expertise and so challenge and enrich the standards used to metacognitively monitor and control writing as it unfolds. Thus, the theory of action provides a fine-grained account of what Sadler terms 'intelligent self-monitoring' (1989, p.143).

The final phase of the theory of action is large-scale adaptation, as per the Winne and Hadwin (1998) model. Large-scale adaptation is the aim of formative assessment practices: making intentional changes to learning in the next episode as a result of reflecting on and evaluating the task just completed. It is through large-scale adaptation that we expect learning to develop as well as outcomes and attainment to improve. However, large-scale adaptation does not always occur (Winne and Hadwin, 1998). Even where writing is explicitly taught, as in the US, expertise does not necessarily carry or transfer from task to task (Donahue, 2012). The present study set up the conditions for large-scale adaptation by focusing on an exam essay genre which writers used over a six month period that culminated in end-of-year exams.

Chapter 4 presents the research design and methodology for the study. The overall approach is design-based research (DBR) methodology, which aims to build and test theory as well as contribute to practice (Brown, 1992; Cobb, Confrey, diSessa, Lehrer, and Schauble, 2003). The thesis reports on one, initial design research iteration (2015); a second small-scale iteration (2016) used slightly different conditions, and there is the possibility of scaling up post-DPhil. Design research studies often start on a small scale (e.g. MacArthur and Lembo, 2009), and the study aims to examine whether the 2015 iteration of the intervention shows promise in supporting self-regulated writing development. The study does not aim to show whether peer assessment 'works,' whether it provides a neat statistical solution to these problems (e.g. Slavin, 2004). Such causal

research questions only belong in the final stages of DBR cycles (Phillips, 2006). The study uses a single-group design, appropriate for this early phase of design research (e.g. MacArthur and Philippakos, 2013), a quasi-experimental design neither ethical nor practical.

Next, the chapter presents the setting and sampling for the research: the first year of the English Language and Literature degree at the University of Oxford. This is an ideal case as the tutorial system offers affordances not present in other HE environments, and students have a clear transition from tutorial writing to the examination essay, the focal genre of the study. The intervention consisted of peer assessment workshops on writing for Prelims Papers 3 and 4, traditional period papers spanning Literature in English 1830–1910 and 1910 – present day. These examinations are sat in June and require three essays written in three hours in response to a set of around 20 unseen items consisting of a quotation with a question. While students write weekly tutorial essays, the examination essays have different conditions and requirements. Marking conventions and examiners' reports are available but student satisfaction survey and Student Union data indicates that these not perceived by students to be sufficient in helping them understand expectations (OUSU, 2015). The chapter then details how I designed the intervention, adapting the principles given by Sadler (2009a, 2010) to the institutional and disciplinary context. In 2015, 21 students took part in four sessions of up to two hours each. Students first trained in holistic assessment, responding to and discussing authentic example essays without pre-determined criteria. The group then wrote a timed essay for anonymous peer feedback, giving and receiving comments from three writers.

To evaluate the intervention and address the research questions, the study employed multiple methods. Ten case study participants met with me pre- and post-intervention for a writing session in which they chose a question from an examination paper and wrote a timed one-hour practice

essay, thinking aloud up until they started writing in full sentences. The essay was followed by a semi-structured interview about the process of writing the essay, as well as beliefs and experiences of writing more broadly. In addition, a range of data was generated during the workshops. Copies were collected of the annotations and responses made on example and peer essays, feedback on peer essays, and of written reflections on learning and writing. The analytical approach combines both data-driven and theory-driven coding in NVivo with case study analysis based on analytical memos and visualisations to map the development of self-regulated writing. The chapter then turns to reflexivity, and in particular combining the roles of researcher and workshop teacher. I outline the ethical procedures followed, and reflect on some ethically important moments that arose during the study (Guillemin and Gillam, 2004). The chapter concludes with reflections on the limitations of the design and methods.

Chapter 5 sets the scene for the findings chapters, introducing headline findings which are then elaborated in later chapters. During the workshops, criteria were not pre-determined, but co-constructed only after participants had responded to and discussed three example essays. The chapter begins by addressing Research Question 1: During the workshops, which criteria do students use to evaluate written essay products? Through an inductive and iterative coding process, managed in NVivo, thirty-four criteria emerged from the data. *Question* was the most frequent criterion by some distance, reflecting its importance for both task perceptions and ensuring task compliance. As the coding structure was flat but many judgements drew on multiple criteria, I then used a cross-tabulation of coding to visualise how criteria interconnect and come together to make up writing quality. *Question* is at the centre of this web of criteria, and the importance of a core of high-frequency criteria surrounding it, such as *clarity* and *argument* becomes apparent throughout the thesis. I then map the student criteria identified in coding onto the Faculty of English Language and Literature marking and classification criteria, which are used

throughout the whole degree course. There is a strong overlap, with only two Faculty criteria of fourteen missing from the coding, one of which is beyond the scope of the workshops ('accuracy of facts'). The Faculty criteria, however, do not mention some high-frequency student criteria, for example *signposting*, *introduction* and *conclusion*. These criteria are important from a student's perspective, and connected with more lofty, difficult to attain criteria. As later chapters show, such criteria may be latent in official representations of writing quality, but play a central role in students' adaptations to writing.

Next, the chapter gives an overview of the adaptations that students made, addressing the over-arching research question, '(how) does students' self-regulation of writing develop in the context of peer assessment writing workshops?'. All ten case study participants adapted their writing over the course of the study, and eight made changes which could be linked to workshop learning. Drawing predominantly on case study analysis, I give an overview of how participants adapted to meet the particular challenges of exam writing, and also how and why the two participants who didn't adapt their writing after the workshops made the decisions not to change further. Finally, the chapter draws on questionnaire and case study data to give an overview of the findings for Research Question 5, 'What is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing?' In the questionnaire data, there was a significant difference between pre- and post-intervention ratings of many items, particularly those relating to understanding writing quality, task compliance and task perceptions. These findings are contextualised with the case study data to unpack how peer assessment provided sources of self-efficacy by making Prelims writing seem more 'achievable', as Megan put it, and allowed students to contextualise their own writing with others' performances.

Chapter 6 further connects peer assessment with self-regulated writing, by examining which criteria become metacognitive standards. The chapter addresses Research Question 2: Which criteria do students use as standards to self-regulate their essay writing processes? Self-regulating writers choose which aspects of their writing to focus on and adapt, and this chapter explores which of the underlying criteria identified in the data participants adopt as standards to guide writing. The chapter opens with some examples of standards in action, and identifies sources of standards for writing: tutorial writing and school examination writing (A-level). While writers see some continuity in between these genres and Prelims writing, they also realise that they need to adapt their approaches. Next, the chapter gives an overview of the criteria coding. There was considerable overlap between the most frequent criteria used to judge example essays, the most frequent criteria used to talk about writing, and employed as standards to regulate writing. However, some criteria which were frequently used to judge others' writing were not employed to regulate writing. It seems that these criteria are too broad to translate into standards, or relate to a reader's overall impression of an essay. Continuing with quantified data, I next compare the group's use of standards at pre- and post-workshops – which standards are brought into emphasis by the experience of appraising others' work. The results indicate that standards related to the question and to task perceptions were more prominent post-workshops, as were standards related to the reader's experience, such as *clarity*.

The chapter then turns to the case study data to examine how Tessa adapted her planning processes in relation to her standards. At the beginning of the study, *clarity* of argument is an ideal for Tessa, transferred from her tutorial essays, but her planning doesn't help her achieve this, and is perceived as 'just sort of… a mess of a random ideas'. The section shows how *clarity* becomes a priority during the workshops. She tries to achieve clarity in her workshops essay by trying a more structured paragraph-by-paragraph approach. However, this is not to her satisfaction. Tessa's

workshop data shows how she pays attention to other ways of achieving clarity, such as setting up the introduction, and her peers offer suggestions for signposting. In her post-intervention essay, a key standard is planning to set up the introduction so that it is 'clear, explicit' and helps her to self-monitor as she writes the rest of the essay. She doesn't outline her paragraphs in order to leave space for 'a more original reading' that really engages with the question. Analysis of other, more routine standards – 'everything that ticks the box' – leads to discussion of the tensions between creativity and consciously trying to meet criteria, a dilemma also voiced by Nicole.

Through the workshops, and as evidenced in how criteria co-occur, students develop a stronger idea of task compliance (Sadler, 2010) – what it means to write an essay that answers the question – and quality. In the theory of action, understanding of task compliance and quality constitute evaluative expertise, and feed into the task perceptions and standards that students generate as they self-regulate writing. Addressing RQ2 and RQ3, **Chapter 7** examines how task perceptions developed for Ben, a case study participant. In the pre-intervention writing session, Ben's task perceptions are quite narrow and quickly defined. He makes the question manageable by focussing tightly on one keyword – 'nature'. His think-aloud protocol shows very brief attention to the question/quotation followed by extensive planning processes mainly involving retrieving knowledge around 'nature'. He feels attending to the quotation in the question is optional – an aid to use 'if you haven't got… your own idea of an argument' rather than a key part of the task. Annotations and responses written during the workshops trace how Ben comes to see that 'for me and my writing, the most important [criterion] is engaging with the Q and the quotation. By setting specific parameters in relation to the question, an essay's argument can become more relevant and more convincing'. He realises engaging with the question and quotation can offer more to quality than just task compliance, and adopts this as a standard. In the post-intervention writing session, there is evidence of large-scale adaptation as Ben uses two strategies: deconstructing the

quotation and defining his thesis. He perceives the task as 'they're trying to get you to show how well you can fit your knowledge to a question'. The think-aloud protocol shows evidence of grappling with the quotation, and transforming as well as retrieving knowledge. The essay itself is more focused on the quotation and the terms of the question, although he departs from some terms, acknowledging that this may be 'high risk' and not an approach to take in the exam itself – his goal for writing in the research situation is different to the 'real' exam. Ben's development of task perceptions and the standard of engaging with the question/quotation is contextualised with data from other participants, including how emotion shapes task choices and perceptions. While the need to engage with the question and quotation may seem obvious knowledge to markers and teachers, it is an example of important tacit knowledge which is hard for students to develop without direct experience.

Chapter 8 examines the impact of *receiving* feedback, addressing Research Question 4: (How) does peer feedback contribute to students' self-regulation of writing? In this study, feedback did not necessarily provide brand new information. Instead, to use Ben's term, students 'amalgamated' feedback with the judgements they had built up on their own writing over the course of the study, using it in a variety of ways. In only two cases did students directly implement peer suggestions, but feedback also served to confirm the success of an adaptation made in the workshops, confirm priorities for further adaptations, and raise awareness of the importance of an aspect of writing. In the latter instance, analysis shows how it wasn't always possible for students to implement suggestions straightaway; rather they had to take intermediate steps in order to reach that point.

The chapter begins with an analysis of the characteristics of the feedback comments, both in terms of criteria and the types of comment given, which predominantly fell into the codes of *praise, problem* or *solution* (building on Patchan, Schunn and Correnti, 2016). This analysis links

with reflections and interview data to indicate that students particularly valued solutions-focussed feedback. Perspectives on praise were more mixed. For some participants, praise helped construct the reader as appreciative and constructive. Positive feedback also provided a 'boost' (Nicole) to self-efficacy for some, helping students see the strengths in their work and realise that 'actually, I can do this' (Felicia). However, a few participants, perhaps more used to robust feedback, were suspicious of praise, or considered the reviewer to appear condescending, so disrupting the symmetrical peer-to-peer relationship. A very small number of reviews also undermined this relationship by using 'blunt' or 'cutting' language (Rachael, Ben). As students received multiple reviews, they were able to contextualise feedback. Feedback, then, perhaps became less important as a catalyst for self-regulated learning, as students could also draw on their own evaluative expertise and metacognitive awareness of writing.

Chapter 9 reviews the findings, revisiting the theory of action and reflecting on the wider implications and contributions of the study. The thesis demonstrates that holistic, comment-only peer assessment shows promise for supporting writing development. Eight of the ten case study writers made adaptations to their writing which could be linked to workshop learning. The value of the theory of action is that it allows the illumination of the different facets and processes of self-regulation that led to such adaptations. The study both exemplifies and extends Sadler's idea of 'intelligent self-monitoring' (1989, p. 143) and shows how peer assessment shapes self-regulation of writing.

A key contribution of the study is that it provides insights into the nature of evaluative expertise and students' own standards. Students' comments drew on a wide range of criteria but showed good congruence with Faculty criteria. Coding the data inductively showed how holistic assessment made visible criteria that are latent in the Faculty criteria. These procedural aspects of

writing were important to adaptation and were part of the shift towards reader-oriented standards. With limited time, students prioritised which adaptations to make to writing, including an increased emphasis on task perceptions and the question, and only two writers directly implemented peer feedback suggestions. Nonetheless, the thesis highlights other ways in which self-regulating learners use feedback, and these findings are contextualised with the wider literature. Peer assessment also impacted on students' self-efficacy for writing, and findings resonated with the broader literature on transition, particularly to élite universities.

As well as theoretical contributions, the thesis also addresses the practical issue of writing support within the discipline. The study shows that Sadler's (2009a, 2010) is proposition is practicable and does serve learning about writing. However, peer assessment is not a standalone writing pedagogy, and should be accompanied by ideas for writing strategies so that students also have the procedural knowledge necessary to develop writing. In summary, peer assessment has the potential to make the transition between school and university smoother and more inclusive, helping students develop their self-regulated learning, or the 'independence' that university teachers cherish and call for.

Chapter 2
Literature Review

2.1 Chapter introduction: the practical problem

It is often assumed that once we have 'learned to write' at school, we should be able to tackle any writing situation, picking up new knowledge from the environment and experience. This assumption underlies the dominant institutional approach to writing in the UK at university level. Only students who are deemed at risk of experiencing writing 'problems' are offered instruction and support: those with disabilities, and those with English as second language (L2) (Hathaway, 2015). Writers who struggle may be offered study skills support from a specialist outside the department (Wingate, 2006). However, since the late 1990s, academic literacies research has raised awareness that academic writing is more than just a set of atomised skills: it also operates 'at the level of epistemology and identities' (Lea and Street, 2000, p.35). Students starting university have to transition to new ways of constructing knowledge. For instance, referencing is often treated as a technical skill and taught as plagiarism prevention. However, knowing how and when to reference is contingent on an epistemological shift towards seeing knowledge as constantly evolving within the field, and needing to build on and challenge what is already thought in one's own writing (Wingate, 2006; Gourlay and Deane, 2012). Lea and Street's (1998) seminal study found that lecturers' use of terms such as 'structure' and 'argument' reflected disciplinary epistemological assumptions, and lecturers found it hard to articulate their own tacit knowledge of such terms. Thus, while on paper there may be a high continuity between A-level and degree English terminology and assessment (e.g. Wilson, Child and Suto's 2016 documentary analysis), students may not experience this. The metalanguage for writing is not transparent and it cannot be taken for granted that new students will share the same understanding (Wingate, 2012, Lillis

and Turner, 2001). As a result, picking up disciplinary conventions can take a long time. For instance, at the end of their first year, many of the psychology students in McCune's (2004) longitudinal study still had very different conceptions of essay writing to their teachers.

A further challenge for new undergraduates is the transition from highly teacher directed learning to an expectation of independent work (Christie, Barron, and D'Annunzio-Green, 2013; Beaumont, O'Doherty, and Shannon, 2011; Burke, 2009). The concept of the regulation of learning is helpful here, bringing together learning and teaching to focus on who manages learning and how (Vermunt and Verloop, 1999). Emerging from régulation in French, according to Allal (2010: 348), regulation 'entails the processes of goal setting, monitoring progress towards the goal, interpretation of feedback from monitoring, adjustment of goal-directed actions and/or of the definition of the goal'. In the literatures, regulation of learning can take many forms, depending on who is guiding or managing the processes. Learning can be self-regulated. If the learner is not yet able to self-regulate, or the situation does not afford or require it, learning can be other-regulated, to use Allal's (2010) term. In the context of school, learning is often teacher regulated, with the level of control varying. Vermunt and Verloop's (1999) work on the interplay between teaching and learning is valuable here. They consider activities such as informing students of learning activities, setting summative exams, and giving feedback to be examples of high teacher control: the teacher assumes responsibility for directing such activities. There is a reciprocal relationship between the level of teacher control and students' level of self-regulation; these need to be well-matched for teaching and learning to be productive. If students are highly self-regulated, then a high level of teacher control may be frustrating and unproductive. Equally, when students arrive at university, they may experience friction if the level of teacher regulation is looser than they are used to, and they are not ready to increase their level of self-regulation in response. Such friction can be constructive, if students are given the tools and support to self-

regulate; after all, a healthy level of challenge fosters self-regulation, and if tasks are too easy, there is no need to take decisions about goals and strategies (Hadwin et al., 2011). However, as Vermunt and Verloop (1999) stress, such friction can also be destructive, impeding learning if learners are left to flounder.

School writing is likely to be often highly teacher regulated, particularly as students reach GCSE (Ofsted, 2009). The National Curriculum in England promotes a 'fossilised view' of writing as involving a linear sequence of planning, writing and revision (Myhill and Jones, 2007, p.339). Teachers may decide when to start, stop and change writing activity, and students may have little opportunity for extended writing. In my previous study, the longest period of writing afforded in 22 lessons with Year 7 – Year 10 students was 25 minutes (Usher, 2013). Powerful measures of accountability, too, lead to high level of teacher control of writing. GCSE and A-level assessment have recently moved away from coursework (Marshall, 2017; Williams and Williams, 2017), limiting extended writing projects. Even before such reforms, students had fewer extended writing opportunities at A-level than at degree, as a comparative survey study found (Green, 2005). It is ironic that the reforms were prompted by a low trust in teacher judgement, and concern that teachers were offering children too much support to meet the necessary grades. In a study at primary level, a shift in emphasis in teaching towards National Literacy Strategy criteria, such as atomised linguistic devices, led to pupils in Grainger, Goouch and Lambrith's (2003) study clamouring for more autonomy and choice, in other words, more opportunities not only for creativity, but to self-regulate. Examination writing becomes so influential that students begin to use GCSE writing scaffolds as young as Year 7 (Ofsted, 2009) or Year 6 (Usher, 2013). A-level assessment objectives shape the teaching of literature (Atherton, 2007), and become the metalanguage used to discuss writing; a metalanguage which continues to influence students' writing at university (Baker, 2017). However, from a very young age, teaching can be structured so

writers can learn to self-regulate writing, as observational research in Canada has shown (Perry, 1998; Peterson and Portier, 2014). In the UK, Robson's (2016) classroom research found that 4 and 5 year olds were actually more likely to self-regulate on their own, as they directed attention to adults when they were present. Many of the strategies used by teachers to foster self-regulated learning (SRL) in these studies, were, in fact, formative assessment strategies, such as peer and self-assessment.

When viewed as a broad pedagogic process rather than a series of events, formative assessment is a key strategy that teachers use to regulate learning processes, not only in planned activities but in classroom interactions (Perrenoud, 1998; Black and Wiliam, 2009). There is growing interest in the intersection between formative assessment and self-regulated learning, in particular (e.g. Nicol and MacFarlane-Dick, 2006; Clark, 2012; Black and Wiliam, 2009). However, as I will argue below, formative assessment is a broad umbrella term and encompasses a range of practices. There is as yet little research into which types of formative assessment best foster disciplinary-specific learning (Yorke, 2003; Bennett, 2011), and writing in particular (e.g. Graham et al. 2011). Designing the most suitable activities is important as Torrance (2012, p.334) argues that 'assessment is always formative, it will always impact on students and have a central place in what and how students learn, but not necessarily in a positive sense'.

Before moving on to the literatures, I wish to address why embedding peer assessment of writing into disciplinary teaching is desirable. As the discussion above shows, since Lea and Street (1998), academic literacies work has raised awareness of the complex challenges students face, and is influential among researchers. However, it has had less influence in the policy arena, and few institutional-level changes have occurred. Academic literacies itself does not yet offer a workable pedagogy (Russell, Lea, Parker, Street and Donahue, 2009; Wingate and Tribble, 2012), with only a

few examples of practice in the literature (e.g. Lea, 2004; Lillis, Harrington, Lea and Mitchell, 2015). A few universities have begun to offer writing centres and courses (Ganobcsik-Williams, 2006; Deane and Ganobcsik-Williams, 2012). However, such initiatives involve considerable resources and will. As research in the US has found, there is also the risk of students struggling to apply writing knowledge from a standalone composition course to disciplinary practices (Donahue, 2012; Wardle, 2007). In addition, a deficit and skills discourse will prevail while writing is delivered outside the discipline by specialists and on an optional basis. Thus, given that knowledge construction and writing are intertwined, it would be best to embed opportunities to learn about writing into the discipline. As Rijlaarsdam, Braaksma, Couzijn, Janssen, Raedts, van Steendam, Toorenaar, and van den Bergh (2008) point out, writers do not automatically learn to write by writing: opportunities to raise awareness of writing are needed for learning. As I will argue in this chapter, the literature indicates that peer assessment is a time-efficient method for such awareness raising.

The chapter begins by presenting a metacognitive definition of writing, and showing how this connects key concepts central to academic writing and to the thesis. I then discuss the current state of play of formative assessment in higher education, and how the preoccupation with teacher feedback rather than student learning is problematic. Next, I examine closely how formative assessment can support writing, first by looking at the problematic notion of 'good' writing and then examining practices reported in the literature, some of which seem problematic for learning about writing. I then look at Sadler's proposal in more detail, and compare two ways in which it can be enacted: exemplar use and peer assessment. The chapter then focusses on peer assessment in more depth, examining theory and research in the writing literatures to examine how peer assessment might support learning about writing, and the research on receiving peer

feedback. The chapter closes by examining key perspectives on the transition to English literature study, the setting for the research.

2.2 Defining writing

As writing is subject to many discourses within and outside of academia (Ivanič, 2004), it is important to begin by defining writing for the purposes of the study. The thesis uses Hacker, Keener, and Kircher's (2009) definition as it highlights processes that are important for the role of formative assessment in writing development. Hacker et al. (2009, p.160) argue that writing processes such as re-reading, editing and revising are all in some way a form of metacognitive monitoring or control, and so the writing process itself '*is* applied metacognitive monitoring and control.' Metacognition was first defined by Flavell (1979, p.906) as 'knowledge and cognition about cognitive phenomena'. Hacker et al.'s (2009) central tenet thus initially positions the definition within the cognitive tradition. Writing researchers often focus exclusively on the cognitive or social or textual aspects of writing (Beck, 2009), despite frequent calls for a 'rapprochement' between the social/contextual and cognitive camps in particular (e.g. Graham, Gillespie, and McKeown, 2013, p.5). However, the authors intend for the definition to have a broader scope:

> 'Writing is the production of thought for oneself or others under the direction of one's goal-directed metacognitive monitoring and control, and the translation of that thought into an external symbolic representation' (Hacker et al., 2009 p.154).

The value of this definition is that although it is centred on the individual writer and how writing is shaped by goals, it recognises the social context as inherent to meaning making and communication. Importantly for this study, the definition also emphasises the agency of the

writer, aligning it with both SRL and formative assessment. Hacker et al. (2009, p.160) argue that the writer's own words on the page or screen are at once both product and process, posing the question 'how can we look at our own writing and not also look at our own thoughts?'. This theorisation of writing positions the student-writer as powerful, privileged: the reader-teacher can never really know the writing process. Inferences can be made from the final product but they are incomplete.

Although the metacognitive nature of writing gives the writer a unique perspective, s/he is necessarily constrained, too, by her/his own standards and perspective. Metacognitive standards are distinct from educational or achievement standards; they are essentially a writer's own beliefs and criteria for learning (Poitras and Lajoie, 2013; Winne and Hadwin, 1998). If a students' self-set standards for essay quality do not align with those valued in the context, the work will not be rewarded in assessment. The writer's own privileged position, seeing the thoughts behind the writing, also makes revision challenging. As a Year 7 student put it in my previous study, 'sometimes the mistakes don't always come out, stand out because it's your own work and you think it's brilliant' (Usher, 2013, p.46). Rijlaarsdam, Couzijn, and Van den Bergh (2004, pp.198-199) point out that the hypotheses about writing that lead us to think that our writing is brilliant (or not) are 'en-passant, implicit, tacit or unconscious... the testing of these hypotheses is rarely conscious or overt'. The role of peer assessment in this study is to open up students' standards and hypotheses for testing. By appraising others' work, students may recognise similar problems in their own, or notice alternative strategies. Thus, peer assessment becomes a central part of learning about writing.

In highlighting the 'goal-directed metacognitive monitoring and control' central to writing, Hacker et al.'s definition also opens up links to self-regulated learning (SRL). Broadly, 'SRL is viewed as

proactive processes that students use to acquire academic skill, such as setting goals, selecting and deploying strategies, and self-monitoring one's effectiveness, rather than as a reactive event that happens to students due to impersonal forces' (Zimmerman, 2008, pp.166–167). SRL is often considered to subsume metacognitive monitoring as one component of models (e.g. Dinsmore, Alexander, and Loughlin, 2008; Winne and Hadwin, 1998). However, while metacognition has theoretical roots that are introspective, SRL also considers the interaction with the environment, and is multi-dimensional, encompassing factors such as motivation, self-efficacy and affect (Dinsmore et al., 2008; Hadwin et al., 2011). Students develop self-regulatory strategies by learning from others (Zimmerman and Kitsantas, 2006; Zimmerman and Risemberg, 1997).

In this study, SRL is key to bringing formative assessment and writing together. In the former literature, there has been much interest in SRL as a way to explain how formative assessment might foster valued types of learning and enable students to become more independent (e.g. Black and Wiliam, 2009; Clark, 2012; Nicol and Macfarlane-Dick, 2006). In the cognitive writing literature, processes such as planning, evaluating and revising are self-regulatory and their frequency correlates with writing quality (de Milliano, van Gelderen, and Sleegers, 2012; Ferrari, Bouffard, and Rainville, 1998). However, whereas SRL places emphasis on the deliberate and proactive, the cognitive writing research tends to be 'agnostic' about whether 'strategies' are consciously enacted (e.g. Beauvais, Olive, and Passerault, 2011; Torrance, Thomas, and Robinson, 2000, p.182). As this is not a cognitive psychology study of borderline conscious processes, SRL is the more useful lens.

A final key aspect to address from Hacker et al.'s definition is the knowledge required to successfully metacognitively monitor and control writing. Writers need declarative, procedural and conditional metacognitive knowledge (Harris, Graham, Brindle and Sandmel, 2009). The academic

writing at the focus of this study is produced for a discourse community of teachers, examiners, and peers all writing within one subject discipline: English Literature. Thus the writing is produced for a shared communicative purpose, and should meet expectations held explicitly or tacitly by the more expert members of the discourse community: the academics reading and assessing the writing. In other words, the writing should use expected genre patterns (Swales, 1990), in, for example, its deployment of content, structure and style. As Hyland (2003, p.24) points out, 'genres help unite the social and the cognitive because they are central to how writers understand, construct, and reproduce their social realities'. Genre knowledge is important to success (Russell et al., 2009; Myhill, 2005). However, genres are not formulaic and static; they are socially situated and vary greatly between discourse communities (Swales, 1990; Berkenkotter and Huckin, 1995). As corpus analysis work has shown, there is no single genre of 'essay', but rather a broad genre family, with sub-genres of essay varying from context to context (Gardner and Nesi, 2012). Not only may students' prior genre knowledge be mismatched with university writing, but they may also lack awareness that they need to adapt, and which genre features are important (Reiff and Bawarshi, 2011). Thus, assessing a range of peer examples may open up students' metacognitive awareness of genre, prompting them to test and alter their hypotheses about writing.

The concepts arising from Hacker at al.'s definition of writing thread throughout the literature review and thesis, and the theoretical underpinnings will be presented in further detail in Chapter 3. The discussion now turns to the role of formative assessment in learning in the extant literature.

2.3 Formative assessment in higher education

This study draws on Sadler's (2009a, 2010) proposal for formative assessment to better support learning in HE by shifting in emphasis from teacher feedback to training students to appraise work holistically. By developing evaluative expertise, students should be in a better position to self-monitor work as they produce it (Sadler, 1989). The value of Sadler's theorising about evaluative expertise is that it re-frames formative assessment as a process that is central to learning and teaching. This is in line with other theorisations of formative assessment such as Black and Wiliam (2009), Allal (2010) and Pryor and Crossouard (2008). Much of the Higher Education literature, on the other hand, tends to focus on formative assessment as an event: feedback. A telling example of this is most citations of Sadler (2010) use the paper to support their discussion of the various problems around feedback practices rather than engage with the proposal.

The preoccupation with feedback is perhaps unsurprising: it can be a highly time-consuming chore for teachers, who understandably want their efforts to be acted upon. As Nicol and MacFarlane-Dick (2006) point out, while teaching-as-transmission is out of vogue, feedback is still transmissive. The interest in the impact of feedback is consistent with the wider field, where substantial reviews have been published (e.g. Hattie and Timperley, 2007; Shute, 2008; Elliott, Baird, Hopfenbeck, Ingram, Thompson, Usher, Zantout, Richardson and Coleman, 2016). However, in the HE literature, there is a problematic tendency for feedback and formative assessment to be conflated. Even Nicol and MacFarlane-Dick (2006) use the terms interchangeably in places. Of course, there has been lots of research attention given to forms of assessment other than teacher feedback, but these are often too classified as 'feedback', for example in Evans' (2013) systematic review. Contemporary higher education conceptualisations of feedback do emphasise the agentic role of learners, and position feedback as dialogic (Nicol, 2013; Carless, Salter, Yang and Lam, 2011; Boud

and Molloy, 2013). Nevertheless, I would argue that using 'feedback' as a blanket term is problematic as in common usage it still connotes an atheoretical, teacher-centred conceptualisation, positioning students as recipients rather than participants who are themselves capable of making judgements about learning. Framing peer assessment as a form of feedback misses the potential learning opportunities for the giver.

Indeed, recipients of feedback do not always see the experience positively. A common theme of peer assessment studies is trust: students do not always have confidence in the judgements or efforts of their assessors (e.g. Cartney, 2010; McConlogue, 2012). In addition, students may expect lecturer feedback to be frequent, directive and focused on how to improve, as they may have experienced at school following the advent of Assessment for Learning (AfL) (Beaumont, O'Doherty, and Shannon, 2011; Price, Handley, Millar, and O'Donovan, 2010). The most problematic issue for academic writing specifically, however, is the mutable and tacit nature of metalanguage used in feedback. As Sadler (2010, p.545) puts it, 'telling can inform and edify only when all the referents – including the meanings and implications of the terms and the structure of the communication – are understood by the students as message recipients.' Terms such as 'argument' may seem transparent and common sense to experienced users of academic discourse, but are problematic for students (Lea and Street, 1998; Lillis, 1997; Wingate, 2012). The idiosyncratic and tacit nature of the hypotheses that a writer holds (Rijlaarsdam et al., 2004) become more problematic when that writer is a lecturer giving feedback, and meanings and values vary between staff as well as students (Carroll, 2002; Read, Francis, and Robson, 2001; Sadler, 2010). Lillis and Turner (2001, p.65) note that language only becomes 'visible' to lecturers when it is a 'problem.' When writing is a 'problem', feedback is likely to be expressed in a more negative and authoritative terms (Court, 2014; Wingate, 2010).

Staff are often well aware of the limitations of feedback (e.g. Price et al., 2010). Integrating peer assessment into courses as Sadler (2009a, 2010) suggests could be a way of integrating academic writing into the discipline without increasing workload. However, some types of peer assessment are more appropriate for learning about writing than others, and this debate is the focus of the next section.

2.4 Formative assessment and writing

Rather than evolving out of specific subjects, the more recent formative assessment literature has emerged from a concern about the impact of assessment on learning (e.g. Black and Wiliam, 1998). The two most prominent theorisations are described as a socio-cultural 'idealised heuristic' (Pryor and Crossouard, 2008, p.17) and a 'rationale' for formative practice that takes in an eclectic range of theory and teacher strategies (Black and Wiliam, 2009, p.5; Wiliam, 2011). However, to truly integrate assessment with learning, theory needs to take account of the nature of the discipline or the domain (Coffey, Hammer, Levin, and Grant, 2011; Bennett, 2011). The principle that ideas of quality are 'clarified' (Black and Wiliam, 2009, p.8) or 'negotiated' (Pryor and Crossouard, 2008, p.15) should be particularly useful for writing as a goal-directed activity (e.g. Hayes and Flower, 1980). However, it is particularly challenging to 'clarify' ideas of quality in writing, and open up teachers' 'guild knowledge' to students (Sadler, 1989, p.126), as this section will show.

Although there is growing interest in formative assessment and writing in the literatures (Hamp-Lyons, 2011), there is as yet little writing-specific formative assessment research (Graham et al., 2011; Hayes, 2004; Nielsen, 2014). Internationally, writing assessment is itself a young field, and only became a discipline distinct from educational measurement in the 1990s, when the journal *Assessing Writing* started (Behizadeh and Engelhard, 2011). In England, L1 writing assessment is

arguably not a distinct field. Writing is subsumed into English at secondary, and there is no equivalent to US compulsory composition courses at HE level. Thus, much of the writing assessment literature cited here is from outside the UK.

2.4.1 Writing quality

As Elbow (2006 p.92) puts it, 'good writing is such a complex construct that there are myriad kinds of 'goodness''. Studies in writing quality and development have traditionally drawn on linguistic measures of quality, as these are measurable (Camp, 2012). There are many ways to achieve a high score on linguistic measures, even in short timed essays in a narrow genre, as Crossley, Roscoe, and McNamara (2014) found. Their computational linguistics analysis of a corpus of essays found multiple distinct linguistic profiles were rated highly on a SAT scale. As writing becomes more sophisticated, it becomes harder to separate components of writing such as 'argument' and 'structure' and even 'language' and treat them as separate criteria that give evidence of distinct types of learning, as Elander et al. (2006) found in their review of HE criteria. With increasing complexity, it also becomes difficult to predict which individual features contribute to quality. For example, it might be expected that a high level of coherence and readability would be associated with higher ratings, but in McNamara, Crossley and McCarthy's (2010) automated analysis of 120 university essays, this was not the case. Essays that scored highly on a SAT rubric were not, in fact, more cohesive, but rather tended to have greater syntactic complexity and lexical diversity. This finding is perhaps explained by the background of the raters: composition teachers, who perhaps value sophistication and require fewer cohesive cues than other readers.

Thus writing quality is socially constructed and context specific. Assessing writing 'does not take place in a vacuum'; assessors necessarily use comparators or referents (Brooks, 2012; p.68). As Wyatt-Smith and Klenowski (2013 p.37) put it, each act of judgment is 'cumulative', referring back to and building on previous assessments. Some college composition departments in the US embrace such divergence, developing assessment criteria locally through discussion and allowing teachers to bring in their own values alongside the core agreed criteria (Dixon and Moxley, 2013; Osborne and Walker, 2014).

Teachers' 'guild knowledge' (Sadler, 1989, p.126), is then, founded in personal experience. Indeed, recent research has found that teachers and academics hold quite strong personal constructs of quality. Sword's (2012) interview study with academics who were considered successful writers by their peers highlighted that conceptions of quality and stylishness varied quite considerably. Lines' (2012) longitudinal interview research mapped six distinct but temporally consistent profiles among secondary English teachers, and many interviewees reported tension between their beliefs about quality and the national curriculum criteria. Their Year 8 students tended to voice similar beliefs about quality, but there was evidence of their own idiosyncratic conceptualisations emerging (Lines, 2014). It is not surprising that teachers' and students' beliefs about writing did not overlap completely. Writing is a creative act, and children develop strong beliefs and attitudes from a young age. Drawing on interview and survey data, Grainger et al. (2003) showed that teachers' conceptualisations of quality were strongly based upon linguistic devices and genre features, aligning with the National Literacy Strategy. Yet their primary pupils showed a strong preference for personal expression and creativity, and wanted more autonomy and choice in writing. Other research has highlighted that a mismatch between teacher and student beliefs about quality can lead to students struggling. Beck's (2006) year-long observational and interview study with one ninth-grade class in the US found that students' success depended on whether

they shared the teacher's goals and were able to enact them in their writing. Similar conclusions have been reached to explain the difference between teacher and student ratings of writing. In Varner, Roscoe and McNamara's (2013) computational linguistics study, tenth-grade students' ratings of their own writing were related to a smaller subset of features than teacher ratings. Limitations of this study include the lack of training students received on rubric use, and the short task duration (25 minutes). Nevertheless, such findings indicate that students need direct experience of assessing writing to challenge their beliefs about writing and develop awareness of why unanticipated aspects of writing matter to quality.

Variability in conceptions of quality can be problematic, then. However, writing assessment will inevitably be full of complexities and difficulties. To support learning about writing, formative assessment should allow students to develop, challenge and refine their conceptualisations of quality. It should introduce students to facets they haven't yet considered, and show them that there is no single route to high quality writing. The next section addresses how formative assessment research has responded to the challenges thrown up by the mutable, context-specific and often personal nature of writing quality, by first examining the wider forces shaping writing assessment.

2.4.2 Measurement influences on formative assessment practices

A key finding of a historical review of writing assessment practice in the US was that measurement theory rather than writing theory has strongly shaped assessment practices (Behizadeh and Engelhard, 2011). This mirrors the parallel but separate development of the wider fields of assessment and learning (Baird, Andrich, Hopfenbeck and Stobart, 2017). Arguably, 'writing

assessment practices still seem to be based on outdated conceptions of writing as a set of skills, mainly focussing on formal aspects of writing' (Behizadeh and Engelhard, 2011, p.206). In the UK, Ivanič's (2004) analysis of different discourses about writing and writing pedagogy found that a skills discourse pre-dominates in the English curriculum, and Lea and Street (1998) argue that this is also the case in HE. Slomp and Fuite (2004) discuss the tussle between reliability and validity concerns in large-scale writing assessment, arguing that both should be addressed, but one is often privileged over the other. In the UK, reliability has arguably won in the case of school writing assessments, for instance in the abolition of the writing component of the National Curriculum Tests in 'English' for Key Stage 2 in favour of indirect measures, or spelling, punctuation and grammar. This use of indirect measures is reminiscent of what Yancey (1999) terms the first wave of writing assessment in US colleges in 1950-1970. She considers that the era's focus on objective tests reflected a time when teaching and classroom assessment practices were brought in line to comply with the reliability concerns of testers. Due to intense accountability pressures, that era has never ended in the UK school system (Torrance, 2018).

The historic preoccupation with reliability can also be traced in some formative assessment practices reported in research. Early reviews of peer and self-assessment research noted an emphasis on the extent of agreement between student and teacher or expert scores (Topping, 1998; Boud and Falchikov, 1989), and this continues as the focus for much contemporary research (e.g. Cassidy, 2007; Kirby and Downs, 2007; Nielsen, 2014). There is often a lack of clarity in articles about the purpose of formative assessment (Gielen, Dochy, Onghena, Struyven, and Smeets, 2011). If the purpose of formative assessment is for students to learn about their learning or writing, and to be able to transfer or generalise this, then surely practices developed for measurement purposes are not appropriate. Reliable scores pinpoint attainment on a particular piece, but do not provide rich information about learning or how to progress further. This section

is thus essentially a validity inquiry for such practices, examining whether they meet the purpose of learning about writing (Stobart, 2012).

The problematic nature of the reliability hangover can be seen in the use of rubrics, marking artefacts that are often used to ensure consistency of marking but also to communicate criteria, and for peer- and self-assessment (e.g. Covill, 2012; Reddy and Andrade, 2010). Rubrics are intended to make assessment more transparent. However, a great number of criteria are salient for writing and so rubrics only sample a selection (Sadler, 1989; 2009b). Rubrics give the impression of analytic judgement based on only the featured criteria. However, other, latent, criteria become salient and manifest as appraisal is not a strictly criterion-by-criterion process: teachers read and appraise the piece holistically, too (Wyatt-Smith and Klenowski, 2013). Thus, as Sadler (2009a, p.52) points out, 'anomalies' arise, for example, but usually remain 'invisible' to students. In addition, some markers may have such strong personal representations of quality that they do not look at the criteria while marking, or even have them to hand (Bloxham, Boyd, and Orr, 2011).

In the Higher Education (HE) formative assessment literature, there are examples of how uses of a narrow range of criteria can encourage students to take a normative approach to writing. For instance, there was some evidence in Bloxham and West's (2007, p.83) study that essay writing became a 'mechanistic exercise' for at least two students from a small interview group. Following an intervention based around criteria, these students reported using the assessment criteria to plan their answers. Similarly, in Andrade and Du's (2007) interview study with college students who had used rubric-referenced self-assessment, participants voiced tensions between following their own expectations versus following their teacher's expectations in their writing. One student expressed this as follows:

> 'We're trained to spew out what the teacher wants but, and that's where, and I'm not sure if this says that we're self-assessing or that we're simply just breaking down what the teacher wants in the paper. Basically you're just giving them what they want … it is self-assessing but what is it self-assessing, it's self-assessing what the teacher wants in the paper' (Andrade and Du, 2007, p.168).

Thus, the rubric excludes the student's own ideas about good writing. This enactment of self-assessment seems to lead to destructive friction between teaching and learning, with students unable to self-assess in the way they value as there is tight teacher regulation (Vermunt and Verloop, 1999). The authors recommend that future studies 'explicitly address the question, 'where is the self in self-assessment, and whose expectations matter?'' (Andrade and Du, 2007, p.169). Rubrics are sometimes co-constructed to take account of group expectations (e.g. Andrade, Du, and Mycek, 2010). There is even an example of a student-negotiated rubric evolving over a semester as the class learns more about writing (Inoue, 2004). However, when rubrics of any origin are used to impose a single standard and lens for writing, the developmental possibilities for writing are restricted (Elbow, 2006). This emphasis on convergence rather than completeness or divergence (Gielen et al., 2011; Miller, 2003; Torrance and Pryor, 1998) is appropriate for ensuring high levels of reliability and also social control, but it is questionable whether it is best placed to help students learn about writing.

A further limitation of relying on a rubric is that the selection of criteria is often limited to the particular learning situation. Just as there is a preoccupation with transfer or generalisability from writing courses in the US (Donahue, 2012; Wardle, 2007), there is a growing interest in formative assessment for lifelong learning in the HE literature. Boud's (2000) term, 'sustainable assessment' is often used: 'assessment that meets the needs of the present without compromising the ability of students to meet their own future learning needs' (p.151). The rather modest ambitions for this definition serves as a reminder that formative assessment is not always beneficial or even benign:

it can even be 'deformative' (Torrance, 2012). SRL has received attention as being central to valued lifelong learning (Nicol and Macfarlane-Dick, 2006). A key defining characteristic of SRL is that is '*proactive'* rather than reactive (Zimmerman, 2008, p.166). Thus, to support SRL, formative assessment should enable students to consider how their own goals and standards align with those that are valued in the community, rather than having these determined for them. I argue that the practices proposed by Sadler (2009a, 2010) have the potential to do this.

2.5 Sadler's proposal

Sadler (2009a, 2010) proposes that all courses should offer opportunities for students to develop skill in complex appraisal. The form of assessment that he proposes is holistic, with only a small number of criteria pre-determined. As criteria make themselves manifest through reading, appraising work becomes an inductive *and* deductive process (Elbow, 2006). Allowing students to 'respond' opens up appraisal of the whole before it is broken down into parts to discuss criteria (Sadler, 2009a, p.57). This gives space for the readerly and interpretative response that a criteria-first approach to assessment can preclude (Usher, 2013). Indeed, the verb 'respond' instead of 'assess' is often used in US college-level composition literature, which places emphasis on personal expression (e.g. Beach and Friedrich, 2006; Evans, 2001; Neal, O'Neill, Schendel, and Huot, 2007). This re-frames writing as an act of communication between writer and reader, rather than an object produced simply for assessment.

By appraising authentic examples of work, Sadler argues that students will develop understanding of three important inter-related concepts: task compliance, quality, and criteria (2010). Task compliance is the extent to which the submission meets the requirements set out by the teacher.

The other concepts are quality and criteria. Sadler describes quality as a unitary construct: 'the degree to which a work comes together as a whole to achieve its intended purpose' (2010, p.544). It is usually easier to 'see' or recognise quality (Sadler, 2002) than to describe it, as Lea and Street (1998) found. Recognising quality is pre-requisite for criteria: as discussed above, latent criteria come into play in different contexts (Sadler, 1985; Wyatt-Smith and Klenowski, 2013). Students need experience to judge when particular criteria are salient and contribute to quality, and when they are less relevant and do not. Such experience should enable students to decide which criteria or areas to prioritise in their own work: to make sense of the noise of possibilities that judging writing throws up.

Sadler (2010) argues that developing understanding of task compliance, quality and criteria will enable students to self-monitor work as they produce it. A further key condition for developing such evaluative expertise is reading a range of tasks and quality. This should enable students to broaden both their repertoire of salient criteria, and of 'possible moves' they could make in their work (Sadler, 2009a, p.50). Sadler's theory is unpicked further in the theory of action chapter, and is contextualised with the SRL theory to examine *how* such complex appraisal may support self-regulated writing. The discussion now turns to the writing and assessment literatures' account of evaluative experiences offered to students, and firstly the ways in which Sadler's proposal could be enacted: exemplars and peer assessment.

2.5.1 Comparing exemplar use and peer assessment

In early work, Sadler (1987) argued that exemplars were helpful for defining quality where achievement standards are fuzzy rather than sharp, as in the case of writing. However, he argues that either a very large number of exemplars need to be provided, or materials such as grade

descriptors need to be provided for new users to understand achievement standards. Using exemplars could be one way of enacting Sadler's proposal, and since 2010, various studies on exemplar use have been published. A large-scale study across ten life sciences courses involved lecturers uploading to a virtual learning environment an unspecified number of exemplars, annotated with examiners' comments (Scoles, Huxham, and McArthur, 2013). Lecturers alerted the students that the exemplars were available but did not use them in class. Simply accessing exemplars on a virtual learning environment (VLE) was associated with better performance in end-of-module exams than not, particularly for students in large classes where there was likely less interaction with lecturers. However, the study didn't capture other factors that could influence performance. In a similar study, Handley and Williams (2012) uploaded two exemplars for students on a large business studies module and required students to respond to self-test questions about quality. They compared the cohort's coursework scores to the previous year, and found no difference. Solely providing an exemplar, then, perhaps isn't sufficient for many students to learn about quality and translate that into coursework. Most students need guided involvement in assessment. Hendry, Bromberger and Armstrong (2011) conducted a quasi-experiment between two campus cohorts taking the same course. Both groups were given VLE access to three exemplars. On one campus, students assessed the exemplars in class, using a marking sheet. On the other campus, students were simply alerted to the location of the materials on the VLE. In focus groups, students reported negative perceptions of the marking sheet, particularly with regard to its ease of use. However, the exemplars were helpful as a guide to writing their own assignments. Students who only had VLE access to the materials didn't use the exemplars as much as a guide for their own work, and found it harder to discern relative quality between the exemplars.

Guided use of exemplars in class is perhaps more likely to have an impact on students' learning, then, as identifying quality in exemplars is challenging. Sadler's (2009a, 2010) proposal also argues for the importance of discussion, again aligning the proposal with the wider formative assessment literature (e.g. Black, 2015). Several research studies have highlighted the importance of dialogue to frame and drive exemplar use (To and Carless, 2016; Hawe, Lightfoot and Dixon, 2017). For example, Carless and Chan's (2017) observational analysis shows how a teacher guided discussion to elicit students' thoughts about quality and managed divergences of opinion.

Some attention has been paid to how students use exemplars to guide their own work. In a more involved study, Bell, Mladenovic and Price (2013) gave first-year accounting students annotated exemplars, grade descriptors and marking criteria for a set of assignments. In an extra element to the previous studies, the authors also set up opportunities for practice and feedback. In written reflective data collected from 49% of the cohort – 119 students – they identified two key profiles for how students perceived and used the exemplars. One group treated the exemplars like a 'recipe' for their own work, and wanted more exemplars to be provided. The other group were happy with 'an idea' of standards and used the exemplars more as a 'guide' or 'framework' (Bell et al., 2013, p.774). It is not clear whether 'recipe' use was problematic, but more attention needs to be paid to *how* students use resources such as exemplars in their own work, lest such assessment becomes 'conformative' (Torrance, 2012). There has been some concern in the literature about the potential for sharing exemplars to lead to plagiarism (Handley and Williams, 2012). However, if students do use exemplars as a model to imitate, or even plagiarise, it is necessary to consider *why* they do so. It may because they don't have appropriate writing strategies of their own to produce the assignment. Also, if few exemplars are shared – such as the two shared by Handley and Williams (2012), they risk being perceived as models or templates to follow. Two exemplars may

also be insufficient for students to feel confident in their own interpretations of quality (Sadler, 1989).

In fact, I would argue that discussing exemplars may help students to develop their ideas of quality, but may not be sufficient or as productive as also then conducting peer assessment. Training is important to peer assessment quality (van Zundert et al., 2010). Using exemplars as training, then having students write their own piece, followed by peer assessment may be more productive. Research from the Netherlands during which writers had the opportunity to write and either read or observe peers' work indicates that the deeper and more demanding engagement with writing quality the better. Rijlaarsdam et al. (2008) designed a series of four lessons with eleven to twelve year old students in the Netherlands, led by Braaksma: the Yummy Yummy case. They first wrote a letter persuading the Yummy Yummy chocolate corporation that they deserved to win a competition. They were then assigned to two conditions: readers and researchers. The readers had to discuss and decide which letters were worthy of the prize. The researchers observed the readers while they were evaluating the letters and decision making, and had to define the underlying criteria used to judge quality. Both groups presented their findings to the rest of the class, before the students had an opportunity to revise their letters. In both conditions students made significant revisions, with an effect size of 0.30 for readers and 1.30 for researchers. Effectively, the readers and researchers were enacting a form of Sadler's (2009a, 2010) proposal: they were holistically assessing letters for overall success, and without predetermined criteria. It can be speculated that the reason why the researchers saw more improvements in their revised texts than the readers is that the task was more demanding: they had to identify and weigh up different criteria rather than just make a decision. This may have enabled them to develop a deeper idea of quality, and also which strategies they should use in their own letter for success.

A further reason why exemplars alone may not be sufficient is that declarative knowledge about writing does not necessarily translate into procedural knowledge that can be enacted (Harris et al., 2009). In the Yummy Yummy case, the students may have recognised procedural strategies for improving writing, as well as developing declarative knowledge about quality. Sadler recognises the declarative-procedural issue in his 1989 paper: 'the ability to evaluate others' or one's own work is not necessarily matched by the ability to produce... the possession of evaluative expertise is a necessary (but not sufficient) condition for improvement' (p.139). Students need the opportunity to practice; ideally, deliberately practising certain aspects of writing (Kellogg, 2008). In the example Sadler (2013) gives in a book aimed at practitioners, the group wrote together, and he included his own piece in the assessment process. So, translating Sadler's proposal into a peer assessment intervention where the pieces are practice essays which don't count for course grades not only gives participants more examples of work to assess, but also gives them opportunities to try out new approaches in a supportive environment.

A further example why peer assessment may be more productive than simply using exemplars is that peer assessment involves writing for an audience. Writing for peers not only reframes writing as an act of communication, but feedback from multiple peers allows students to see what different readers make of their work. Writing becomes a more authentic task. Much of the writing students do in wider HE is for summative assessment, and doesn't get an immediate reading (Catt and Gregory, 2006). While students at Oxford have frequent opportunities to discuss essays with their lecturers and peers in tutorials, they receive no feedback on work submitted for examination. Thus, peer assessment of the examination genre can make the writing task more authentic, reminding students that they are writing for a community. In fact, the aim of peer assessment in some settings is to build in more writing opportunities to courses with large enrolments, as Schunn (2016) describes for the SWoRD online platform. An interesting finding from a study using

SWoRD with Physics undergraduates in the US is that draft lab reports written for peer assessment were of higher quality than drafts written for teaching assistants (Cohen's d = 0.5) (Patchan, Schunn, and Clark, 2011). Even though peer assessment was anonymous, writing for an audience of multiple peers perhaps made writing more meaningful and public, and so motivating. In addition, training students in peer assessment could also give them the skills to do so outside the taught curriculum. With their peers, students develop their own culture of learning within the university (Havnes, 2008). Orsmond, Merry and Callaghan (2013) describe how high-achieving students form communities of practice, developing their own self-assessment skill through informal practices and discussion with peers.

The next section examines research on peer assessment and writing. First, I will give an overview of the broad, expanding field. I will then look at findings specifically for writing, and turn to the writing literatures to address the question: 'how might students learn from peer assessment in writing?'

2.6 Peer assessment and writing

Research interest in peer assessment has been growing in recent decades, and particularly since Topping's (1998) review on peer assessment with HE learners, which found considerable variation in practices. Such diversity continues (Gielen, Dochy and Onghena, 2011; Adachi, Tai and Dawson, 2017). The field is described by Kollar and Fischer (2010, p.345) as in a phase of 'adolescence', with a key model of peer assessment processes as well as the field's relationship to other fields yet to be defined. The research specifically on student writing is more limited. Much research in peer assessment has been conducted with L2 learners, and while similar themes are reflected in that

research (Yu and Lee, 2016), unless specified, only L1 research is cited as the challenges of L1 and L2 writing are different. Particular trends in L1 peer assessment on writing include the comparability of peer, instructor and independent rater scores on a rubric (Moxley and Eubanks, 2016) and comments (Cho, Shunn and Charney, 2006; Anson and Anson, 2017). The potential of online platforms such as has also been examined (Moxley, 2013; Schunn, 2012), and the possibility of substituting teacher assessment for peer assessment in Massive Open Online Courses (MOOCs) (Meek, Blakemore and Marks, 2017). Much teacher research in Higher Education has found positive benefits for peer assessment on drafts (e.g. Ludemann and McMakin, 2014; Sampson and Phelps Walker, 2012). In the wider HE literature, a large number of papers have examined student perceptions of peer assessment (e.g. Liu and Carless, 2006; Cartney, 2010; Ludemann and McMakin, 2014).

Van Zundert et al. (2010) note in their review of peer assessment more widely that few studies separate out the impact of giving and of receiving peer feedback. For writing, a series of intervention studies with both experimental and qualitative designs have found that reviewing peers' work was beneficial to students' own. In a four-week study using the online SWoRD system, Cho and Cho (2011) asked 72 Physics students to write and upload draft lab reports. They then reviewed 3 or 4 peer-written drafts, using a rubric to generate both scores and comments on strengths and weaknesses. Once they received the comments, they revised their drafts, re-submitted them to the system and rated the quality of the reviews they received. These were then made visible to the reviewers. Finally, the revised lab reports were re-reviewed. Overall, Cho and Cho (2011) found through regression analysis that the quality of the comments a participant received did not have a significant relationship with the quality of their revised text. However, the quality of the comments participants gave on others' writing did have a significant impact on the quality of their revised lab report. It seems this effect was driven primarily by two types of

comments: those related to the strength of macro-features such as organisation across paragraphs, and those on weaknesses within paragraphs. Perhaps unsurprisingly, making surface comments such as highlighting spelling, punctuation and grammar, did not have an effect.

A further study with Physics undergraduates used an experimental design to compare the effects of giving peer feedback with similar activities. In Cho and MacArthur (2011), before being assigned to a condition, all participants trained to use a rubric and scored two examples, without making comments. The examples were introductory sections of lab reports. The participants were then randomly assigned to a condition. Reviewers read, rated and commented on three example essays; readers read over and over the same example essays, but could not make comments or rate the essays. The no-treatment group simply read articles on a different topic. After an hour, the students then wrote their own introduction to a lab report, on a different subject. Writing quality was scored with the same rubric used by the students. Cho and MacArthur (2011) found that the reviewer condition, i.e. rating and commenting on exemplars was associated with an effect size of 1.28 compared with the no-treatment condition. In addition, there was an effect size of 0.98 compared to reading the exemplars and not scoring or offering feedback. This suggests that participants were able to learn about the lab report introduction and apply that learning more effectively through reviewing: the authors highlight that all students had access to the criteria when writing. However, control conditions in experiments need to be meaningful (Graham and Harris, 2014). A limitation of this study then, is that the non-reviewer conditions arguably were not: reading example introductions or articles over and over for an hour could be frustrating and demotivating for students.

Such findings for the promise of learning through reviewing are mirrored in two qualitative studies in which participants have viewed giving feedback more beneficial than receiving feedback from

peers (McConlogue, 2014; Nicol, Thomson, and Breslin, 2014). This may partly have been due to students in both studies holding low opinions of the quality of feedback offered by peers, a trend seen elsewhere in the literature (e.g. Cartney, 2010). These two studies also give some insight into the participants' judgement processes. McConlogue (2014) followed one student who participated in a summative peer review intervention with Engineering students (reported in McConlogue, 2012). This single-participant case study showed how the student gained insights into quality, but put an unusually high amount of time into writing peer feedback: 4-5 hours to assess 4 papers. She was disappointed with the feedback received in return as the comments were not easy to understand and didn't show her how to improve her work. The findings from focus groups and interviews with 15 students in Nicol et al. (2014) are more broad. While students reported using the criteria questions provided by tutors to guide their judgements, the authors argue that there is evidence of 'a background reflective process involving an active comparison of the other's work with their own' (Nicol et al., 2014, p.115). The authors also speculate, with little support from the data, about how students moved between a holistic process and a more analytical, criteria-supported process when making their judgements. It should be noted that a limitation of both of these qualitative studies is that they are under-theorised. For instance throughout Nicol et al. (2014) there is a very broad, sweeping use of the term 'self-regulated learners' without much indication of what is meant.

While these studies show promise for the impact of peer assessment on writing, existing research has not addressed *how* students learn from peer assessment in much detail. The following section draws on the writing literatures to contextualise why peer assessment may be helpful for learning about writing, beginning with the opportunity to take a readerly perspective on writing.

2.7 How might students learn from peer assessment of writing?

2.7.1 Taking a readerly perspective

Assessing peer texts gives writers the opportunity to texts similar to their own, experiencing interesting ideas that are easy to interpret, or the frustration and difficulty of decoding the writer's intent. Learning to anticipate a reader's representation of one's own text is highly important to writing development (Kellogg, 2008). As writers develop in expertise, their processes and products become more reader-oriented. However, this is a slow process as writing is such a highly cognitively demanding activity (Graham, 2006). Beginning writers' processes are typified by knowledge-telling – retrieving and recording what the writer remembers about material or the story (Bereiter and Scardamalia, 1987). This mode of writing involves no metacognitive monitoring or control (Bereiter and Scardamalia 1987; Graham et al., 1998). Children then move towards knowledge-transforming, which involves processes such as planning, translating and reviewing, deployed more strategically. Writers begin to revise and use other strategies, but the focus is still on the author's representation – how well the author feels s/he has translated her/his ideas onto the page. Kellogg (2006, 2008) proposes a third stage: knowledge-crafting. At this expert stage of writing development, the writer is able to hold in working memory the representation a reader would make of the text, alongside the writer's own intentions, and how s/he represents the text. Transitioning to knowledge-crafting is a considerable developmental leap. While school writing pedagogy addresses audience awareness, 'being aware of a fictional reader in generating text is different from being able to read the text as it is currently written from another person's perspective' (Kellogg, 2008, p.9). Offering students the opportunity to assess peer texts may help students take a readerly perspective on their own writing by noticing similar issues in others' work.

While Kellogg (2008) argues that knowledge-crafting requires the writer to have automated many processes to free up working memory, typically after twenty years' experience, other factors may too play an important role. Lindgren et al. (2011) argue that use of strategies and genre knowledge are important to expert writing. In an intensive study with six participants, they used interviews, stimulated recall and keystroke logging to compare ten-year-old, fourteen-year-old and adult professionals' approaches to writing tasks. While the ten-year-old writers predominantly used strategies associated with knowledge-telling, the distinction between the teenage knowledge-transformers and professional knowledge-crafters was less stark. One fourteen-year-old had a similar level of reader awareness to the adults and frequently made revisions to anticipate the reader's needs. However, the adult writers had a broader repertoire of strategies and genre knowledge, which they deployed intentionally; the teenage writers' awareness was more implicit. Thus, the ability to hold possible reader representations in working memory is not the only hallmark of knowledge-crafting: genre knowledge and metacognitive strategies may also be important.

As Lindgren et al. (2011) found, knowledge-crafting writers make revisions to anticipate their reader's needs. Revision is a highly challenging process, and novice writers tend to make limited and superficial revisions (Rijlaarsdam, Couzjin and van den Bergh, 2004). Myhill and Jones (2007) found that Year 9 and Year 11 students made various types of revisions, including for coherence and repetition, but tended to see revision as an end of task activity, to check over what they had written; few re-read their work in full. Hayes (2004) suggests that writers use different modes of reading as they revise. These modes of reading depend on writers' goals and may explain some differences between novice and expert writers. Teachers, for example, may tend to read to evaluate as default, drawing on a wide knowledge of quality, including genre. On the other hand, less experienced writers may read for understanding or to check spelling, missing potential

revisions opportunities they have the skills to enact. Developing writers may need to redefine their reading-to-revise goals, then. In a classic experiment, Wallace and Hayes (1991, p.58) asked undergraduates to revise a short text to make it 'clear, organised, easy to read and free of errors'. While the control group received only written instructions, the treatment group received an eight-minute intervention during which the researcher modelled the differences between novice and expert approaches to revision, with the latter including global as well as local revisions. The treatment group made more revisions and improved the quality of the text much more. This suggests that redefining the revision task as a global revision task was influential. In follow-up studies, the findings were replicated with freshmen writers, but not with 'basic writers' starting community college with a lower level of literacy, whom the authors argue may not yet have detection and other skills necessary for revision (Wallace, Hayes, Hatch, Miller, Moser and Silk, 1996).

As well as seeing revision as a global rather than local process, writers need to be able to draw on knowledge about quality in writing to detect errors and see alternative rhetorical possibilities. It is not possible to revise without first being able to detect problems (Hayes, 2004). The importance of knowledge about quality and genre is highlighted in a study of working memory and revision with first-year psychology undergraduates. Adams et al. (2010) asked participants to identify and revise errors in short essays, synthesised from past exam scripts. The essays contained 18 language, structure and argument errors in total, and on average, participants correctly identified 5.79 (SD = 2.27) and diagnosed 4.32 errors (SD = 1.91). Revision was not associated with working memory, nor topic knowledge, although there was correlation with verbal working memory. Given wider research findings on the varied and contested meanings of terms such as 'argument' (e.g. Lea and Street, 1998; Wingate, 2012), it is unsurprising that untrained first-year students had difficulty recognising and correcting what staff members consider constitute 'errors' of argument or

structure. Indeed, the participants reported that they found the task very difficult. Even the researchers had difficulty agreeing on errors, as correlations between rater scores for argument showed. Judging whether participants had diagnosed an argument error yielded .20 and providing an adequate revision .36. These findings point to the importance of developing shared ideas about quality, as well as genre knowledge: it is unlikely the participants had experience of the exam genre prior to the task. This leads to the next area in which peer assessment may support writing development: genre knowledge.

2.7.2 Learning about genre

Genre knowledge is important not only for detecting problems in a task but for guiding writing in the first place. Writers can learn from genre about reading, but the mode of reading is again important for the quality of learning. In a series of experiments, Bereiter and Scardamalia (1984) found primary, secondary and postgraduate students were all able to notice some features of genres from a single model text, and apply a few in their own writing. However, at all levels students tended to identify and use word-level and content-based features rather than global aspects of writing. For academic writing, a more sustained approach, genre analysis of model texts, is a widely practised in English for Academic Purposes courses for L2 learners (Swales, 1990); with some similarities to the systematic functional linguistics approach in Australia (Flowerdew, 2015; Bawarshi and Reiff, 2011). Typical features involve unpicking model texts to identify the rhetorical moves that writers make, and the macro structure. Even a relatively short intervention can raise genre awareness. Over a twelve hour course during which they analysed six texts, Negretti and Kuteeva (2011) found that L2 students were able to develop their declarative and procedural knowledge of research article writing. However, the course did not require students to

write in the research article genre. Other studies have found that genre awareness does not necessarily lead to production. Also, writers may unintentionally reproduce features of prior genres when writing in new ones. In Bereiter and Scardamalia's (1984) study, there was little difference between children and adults' learning about and replication of a made-up genre, 'concrete fiction'. While adults participants noticed that the model was written from an objective perspective and included detailed physical description, they often did not include such features in their own piece; rather they tended to reproduce more standard fictional genres. These findings are mirrored in more recent research. Artemeva and Fox (2010) asked engineering students to name the genres used in a range of passages, and then write a technical report. While around 80% successfully identified one of the passages as a technical report, only 11% then wrote a text which raters considered to be a technical report. The majority wrote essays instead. Those who used the technical report genre had had exposure to the genre on work experience, suggesting that experience is necessary for successful production.

Writers' unwitting use of prior genre knowledge suggests that students need opportunities to develop awareness of which aspects of prior genre knowledge can be transferred to new tasks, and which cannot. Reiff and Bawarshi (2011) interviewed students at two US universities about the texts produced for a first year composition course. They identified two types of talk about writing and transfer from prior genre knowledge, drawing on Perkins and Salomon's (1988, 1989) low-road and high-road transfer. Boundary guarders were very confident in their talk about writing, but tended to replicate whole genres, even where inappropriate, demonstrating low-road transfer. Boundary crossers were more prepared to adapt their knowledge, talking more tentatively about writing, and singling out genre strategies that they could apply to new tasks, reflecting high-road, or deliberate transfer. Even where students are boundary crossers, ready and willing to adapt their knowledge, they may not have the opportunities to apply their writing knowledge. In a further US

study, Wardle (2007) followed seven writers from first-year composition into subject teaching over two years. Although the writers had developed what she terms 'meta-awareness' in her course, they rarely had the opportunity to apply composition strategies as subject assignments were either insufficiently challenging, or did not allow time for drafting and re-drafting. To develop their writing, students need genre knowledge, awareness of which strategies they can apply, but also opportunities within the curriculum to practise writing meaningfully. Kellogg (2008) suggests that deliberate practice of aspects of writing is how writers become knowledge crafters.

For students in the UK, where writing instruction for L1 undergraduates is rarely provided, school genre knowledge and conceptions about writing may continue to be influential. The little research on writers' transition from school to university indicates that this is the case. Baker (2017) conducted ethnographic research with a small group of writers as they transitioned from school to university. She found that an assessment discourse of writing influenced students, with one writer using A-level assessment objectives to guide her writing. Perhaps this should not be surprising. Indeed, in terminology alone, A-level is designed to prepare students for degree, and in course documentation alone, assessment types and terminology are fairly similar, as Wilson et al.'s (2016) analysis shows. However, as corpus research has shown, 'essay' is a genre family, with different genre variations between contexts (Gardner and Nesi, 2012). Students need, then, to see real examples of genre in different contexts. While genre instruction would be ideal, this may not be practicable within the discipline. Peer assessment can provide such a similar opportunity, particularly if students appraise a range of examples, as proposed by Sadler (2009a).

2.7.3 Observing others' performances

While genre pedagogies have traditionally used models of writing, Devitt (2015) argues that such approaches should now move towards performances or examples of genre in use, and how these vary. Similarly, research in the social cognitive tradition of SRL suggests that students learn more from imperfect, coping models than from models who already show mastery of a task, and complete it with ease. Schunk and Zimmerman (1997) and Zimmerman and Kitsantas (2006) posit that observation is a first step in learning a new skill. In an experiment, Zimmerman and Kitsantas (2002) asked female high school students to combine sentences. Students were allocated to three conditions. In the first condition, there was no model. In the second condition, students observed a model practising the sentence combining task, often making mistakes but gradually learning how to self-correct: as a coping model. In the final condition, the students observed a mastery model, who completed the sentence combining exercises without making mistakes. Observing either model was more effective than not, and students in the coping model condition gained the highest scores in the final sentence combining task. Zimmerman and Kitsantas (2002) argue that such vicarious experiences help students develop strategies and make adaptations when learning does not go as planned.

Raedts, Rijlaarsdam, van Waes and Daems (2007) asked undergraduates to perform a more complex and ecologically valid task: writing a synthesis of research articles. In the experimental study, the control group prepared for writing by carrying out some exercises. The experimental group prepared by observing videos of a pair of models thinking aloud as they completed one of the exercises. Participants in this condition were also asked to outline the different steps the model took to complete the exercise, identify which model was more effective at the task, and

explain why. The experimental group outperformed the control group on a number of measures of writing quality in their final writing task.

These studies, suggest then, that writers may benefit from observing others similar to them: the similarity hypothesis, as Rijlaarsdam et al. (2008, p.68) term it. In peer assessment, students do not have the opportunity to observe the writer in the act of composition, but essays provide traces of process – for instance, if the writer ran out of time, and also textual strategies, for example genre moves. Observing others' performances is also an important source of self-efficacy, the belief that one can do a task, as it allows the student to reappraise her/his own performances in the light of others', perhaps seeing it as a mastery experience (Pajares, 2008). Self-efficacy is important to self-regulated learning: if students do not feel they can perform well, they may not use strategies to manage the learning process (Schunk and Usher, 2011; Lavelle, 2009). Formative assessment is considered to be helpful to self-efficacy as it allows students to observe others' performances. However, as yet, there is little empirical research in this area for writing. Hawe et al.'s (2017) case study of one class of business studies undergraduates found that self-efficacy was a key outcome from using a range of exemplars throughout the term: students felt that the task was more achievable. There are also mixed results in the self-assessment literature for the impact of self-efficacy. Andrade, Wang, Du and Akawi (2009), working with middle school writers, found that rubric-referenced self-assessment compared to informal reflection improved girls' but not boys' self-reports of self-efficacy. Panadero, Alonso-Tapia and Reche (2013) asked education students to work with either self-assessment scripts or a rubric over a term as they designed multimedia resources as a WebQuest. They found no significant difference in self-efficacy between the scripts and the rubrics, although the rubrics were perceived more favourably by students. Similarly, Covill (2012) found no difference between conditions on self-efficacy for three types of self-assessment: open self-assessment, self-assessment with a long rubric and self-assessment

with a short rubric. Unlike the other two studies, Covill also looked at the impact on task performance, but found no difference in writing quality between the three groups' Psychology assignments. Again, Covill's participants perceived rubrics more positively. While rubrics are seen as user-friendly, then, they do not seem to impact upon other forms of self-assessment, or no self-assessment at all (Andrade et al., 2009). A commonality in the procedure of all three studies was that students used one example (Covill; Andrade et al.) or even no example (Panadero et al.) to train to self-assess. So, it can be argued that students need to observe others' performances in some way to develop their own idea of standards. Simply self-assessing will not have an impact on self-efficacy if students cannot see others' performances with which to compare their own. Thus, developing ideas of quality is important not just for students to be able to self-regulate writing, but for self-efficacy beliefs.

2.8 Receiving peer feedback

The sections above have explored possible opportunities for learning by giving feedback. In peer assessment situations, we might also expect students to learn by receiving feedback, but as this section discusses, this is a complex and messy process, which has not been sufficiently explored in the literature. As Higgins, Hartley and Skelton (2001, p.273) note, feedback is 'an essentially *problematic* form of communication involving particular social relationships', throwing up issues such as power and emotion. Higgins et al. (2001, p.273) also called for research on what they term the 'inside-out' perspective, or the learner's perspective. Yet surprisingly little assessment research has considered how learners interpret and use feedback of any source. As a recent review by Winstone, Nash, Parker and Rowntree (2017) highlights, attention has tended to focus on the message, and how this is constructed, rather than the recipient. This is unsurprising, given

that is how feedback is conceptualised in student satisfaction surveys: their policy clout has led to an emphasis on educating teachers to message feedback better (Nash and Winstone, 2017).

Feedback is increasingly theorised in the literature as involving dialogue between students and others, with the teacher creating the conditions for learning through feedback, rather than dispensing messages and edicts (e.g. Nicol, 2010; Boud and Molloy, 2013). Such conceptualisations have been influenced by the self-regulated learning research, which emphasises that peers and teachers are not learners' only sources of feedback. As they write or study, learners form their own judgements and self-generate feedback, adjusting their goals or strategies if such feedback reveals it to be necessary (Butler and Winne, 1995; Winne and Hadwin, 1998). Such 'internal feedback' or metacognitive monitoring is gaining some acknowledgement in the feedback literature (Butler and Winne, 1995 p.246; Nicol and MacFarlane-Dick, 2006; Nicol, 2013). It is partly acknowledged in the definition used in Hattie and Timperley's (2007, p.81) widely cited review on feedback: 'information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance'. However, there is less acknowledgement that as Butler and Winne (1995) emphasise, feedback from 'agents' such as peers is filtered or mediated by the recipients' own beliefs about learning, their knowledge, and indeed the conclusions of their own self-generated feedback, or metacognitive monitoring. Recent research has examined how self-theories shape receptiveness to feedback (Forsythe and Johnson, 2017), and the transition literatures show that students' experiences at school are also influential (e.g. Beaumont et al., 2011). However, there is little research on learners' own metacognition and how this comes into interaction with feedback from peers and others. This is discussed in more depth in the next chapter with respect to the self-set, personal metacognitive standards that students use to monitor their work. For the moment, the discussion focuses on the assessment and writing literatures' dealings of how learners receive and use peer feedback.

In order to encourage uptake and embed feedback as a cycle in the writing process, some studies ask participants to submit revisions after peer feedback. Three studies using this mode of peer assessment also analysed the extent to which students implemented comments. The rate of implementation in all three was low: in history courses in the Netherlands, writers implemented around one third of comments (Van den Berg, Admiraal and Pilot, 2006), and in Open University wiki assignments, students addressed almost half of comments on 'shortcomings' (Walker, 2015, p.242). In a larger study with Psychology students in the US and the SWoRD platform, students again implemented only a third of actionable comments (Patchan, Schunn and Correnti, 2016). In the last study, Patchan et al. (2016) examined the features of feedback, and amongst a multitude of analyses found only two factors which made it more likely for writers to action comments: where the location of the issue was given and feedback contained more praise than average. However, if praise was used to soften the impact of critique, it hindered implementation. Interestingly, comments that offered direct solutions were not likely to be implemented, contradicting a previous study, where solutions were more likely to be enacted (Nelson and Schunn, 2009). While it is clear that students are selective about what they implement, these studies do not shed much light into why this is the case. That the SWoRD studies found few clear patterns among many variables indicates that individual differences between students are at play, and there is a need for detailed and qualitative accounts of how students interpret and decide to use feedback.

The few studies that do focus on recipients' processes offer some insights. In interviews with 15 students, Kaufman and Schunn (2011, p.402) found that despite perceiving peer assessment negatively, students who 'made the most changes talked in a more reflective way about how the SWoRD process guided improvements and revisions to their paper'. It seems that those who took value from the experience were already disposed to do so: this describes productively self-

regulating learners who were able to integrate insights peer assessment into their own self-monitoring. As these findings imply and research in the wider field confirms, students may not have good strategies to use feedback (Weaver, 2006; Burke, 2009; Winstone, Nash, Rowntree and Parker, 2017).

A further tranche of research has found that the credibility and perceived competence of the reviewer may affect willingness to act on peer comments. This has arisen in research as diverse as focus group research with undergraduates on lecturer feedback (Poulos and Mohony, 2008), and experimental research with peer assessment scenarios (Strijbos, Narciss and Dünnebier, 2010; Berndt, Strijbos and Fischer, 2018). Indeed, peer feedback can be perceived negatively by students where it substitutes for actions which are usually carried out by a teacher, and particularly where peer assessment takes a summative function, counting towards the degree (McConlogue, 2012; Patton, 2012). In the questionnaire element of Kaufman and Schunn's (2011) study, students in classes where peers alone allocated grades rated the SWoRD system more negatively than students in classes with peer and teacher grading. Power relations are inherent in assessment (Boud, 1995), and students find it disconcerting for peers to be awarded such power (Patton, 2012). Trust is important to any relationship involving assessment (Carless, 2013), and where peer assessment disrupts parity between peers, it seems learning is also disrupted.

Peer assessment, then, can be an emotionally unsettling experience, like any assessment (Boud and Falchikov, 2007). Receiving any feedback is an intrinsically emotional process, and can be especially challenging where feedback clashes with learner's own judgements of learning (Molloy, Borrel-Carrió and Epstein, 2013). Getting a lower grade than expected can invoke strong negative emotions (Ryan and Henderson, 2017), and colour students' interpretations of the feedback itself, and so its uptake (Pitt and Norton, 2017). Learning is not always a pleasant experience, and

difficult feedback in itself is not necessarily a bad thing. However, students can develop defensive habits, which can inhibit learning (Molloy et al., 2013). With regard to peer assessment specifically, there is little research that goes beyond acknowledgement of the emotional. Panadero's (2016) systematic review on the inter-personal and emotional aspects of peer assessment identified 26 articles, but among them, 10 distinct themes. This indicates that as yet the sub-field is emergent and the picture rather disjointed.

Nonetheless, recognising that peer assessment involves emotion, studies often take steps to mitigate the risk of discomfort, or at least control for it. As discussed above, to ensure there is trust in peer feedback and that it is perceived as a learning activity, it may be important to remove any summative function. Training is important both to the quality of peer feedback and to ensure trust in this quality (van Zundert et al, 2010). Training is also important as students may lack confidence in their own ability to give feedback (Hanrahan and Isaacs, 2001; Cartney, 2010). To reduce discomfort and/or control for inter-personal bias, many studies use anonymity (Panadero, 2016), although this can create an administrative burden (Ballantyne, Hughes and Mylonas, 2002). Topping's (1998, p.267) review noted that peer assessment studies tended to conduct grading as anonymous and 'distant', but qualitative comments involved 'personal contact', reflecting their 'personalised' nature. However, as the majority of peer assessment studies are now to be conducted online, anonymity is more usual (e.g. Schunn, 2016; Huisman, Saab, van Driel and van den Broek, 2018). Granting anonymity may enable reviewers to provide more comments, of which a higher percentage may be critical, as Howard, Barrett and Frick (2010) found. In addition, using more than one peer reviewer may enable writers to contextualise comments and weigh up their validity. If feedback converges on the same recommendations, it may be more persuasive (Schunn, 2016). The particular choices taken in the design of this study will be discussed in Section 4.5.

2.9 Transitioning to university writing in English Literature

The final section of this chapter briefly addresses the literatures on the setting for the study: the transition from school to university. This chapter has already highlighted some key themes of this tranche of research. Firstly, Section 2.1 discussed the shift from tight teacher regulation to more self-regulation (Beaumont et al., 2011; Christie et al., 2016), as well as the enduring influence of school assessment practices on writing (Baker, 2017). Secondly, Section 2.4.1 examined the challenges posed by contested nature of writing quality (Gourlay, 2009). It is worth noting that there is little research on students' experiences of transition in the subject of English. Studies tend to focus on comparing perceptions of school and university English teaching, with little reference to writing (Ballinger, 2003; Green, 2005; Smith and Hopkins, 2005; Atherton, 2007; Snapper, 2007). Nonetheless, in Alder's (2018) phenomenographic interview study with undergraduates at a Scottish university, writing emerged as a source of stress for English students, particularly because of uncertainty around conventions. One first-year found each of her three lecturers communicated different expectations for coursework. This finding resonates with Lines' (2012) research on secondary English teachers' beliefs about writing quality, discussed above in section 2.4.1. At degree level, writing quality is likely to be particularly contested because of the nature of English as a discipline, as the discipline itself is fragmented and multiple (Martin, 2006; Chambers and Gregory, 2006).

While in the US, longitudinal research has traced how students' written texts develop over the course of a degree (e.g. Sommers and Saltz, 2004), studies in the UK have focused on learners' development more holistically. A particular focus of the academic literacies research has been learner identity, for instance in becoming an English student (Alder, 2018) or developing academic identity more broadly (Tapp, 2013). Much has been detailed about writing and learner identity

(Lillis, 2001; Lea and Street, 1998), and this research also highlights how the first year experience is emotionally demanding (Christie et al., 2008). While assessed writing can be a source of such struggle, it is also significant for building a sense of identity and belonging. For instance, Gourlay (2009) found that receiving positive feedback or grades marked turning points during the first year, boosting confidence and so sense of identity as a successful student. Reay, Crozier and Clayton's (2009) ethnographic research with working-class students at an élite university is of particular note for this study. The authors characterise students as having developed resilience and self-regulatory skill at school, but describe how when faced with 'the shock of the elite' on entering the university 'bubble', questioned their academic ability and deserving of a place (Reay et al., 2009, p.1111). Such self-doubt was finally challenged by success in first-year examinations. There are resonances here with the self-efficacy research discussed above (Section 2.7.3), with examinations finally allowing students to observe others' performances, however briefly, and so contextualise their own. While learner identity is beyond the scope of the study, the next chapter will consider the role of self-efficacy and emotion in the self-regulation of writing in more detail. The particularities of Oxford as an elite university are discussed later (Section 4.3).

2.10 Chapter conclusion

This chapter has explored diverse literatures to examine the promise for peer assessment to support writing. The analysis has drawn on research on exemplars, rubrics, the role of readers in writing, genre, observing others and receiving peer feedback. Both theory and empirical research indicates that there is much promise for peer assessment to support learning, but the processes and mechanisms have not yet been studied in much detail. A richer understanding of whether and how students learn through peer assessment would inform both practice and theory, and could

make a real difference to students experiencing the challenging transition to university writing. The research questions for this thesis, presented in the next chapter, are intended to address this important research gap.

Analysis of the literature also leads to principles for peer assessment that is tailored to the complexities of academic writing in English literature, and in particular the thorny problem defining and sharing definitions of writing quality. Drawing on the literature, the procedure for the intervention was designed as follows. Students appraised holistically four example essays before writing their own. They then assessed three peer works anonymously, and received feedback from four peers. Finally, they appraised and discussed one final example, before discussing strategies for developing their writing. So, peer assessment involved thorough training, no pre-determined criteria, was conducted anonymously, was comment-only, and provided a practice opportunity as well as discussion of strategies. Next, Chapter 3 outlines the theory of action that describes how participating in such peer assessment should enable students to self-regulate their writing more productively and presents the research questions.

Chapter 3
Theory of action

3.1 Introduction

In the previous chapter, the literature review showed that there is growing research evidence for the beneficial impact of giving peer feedback on writing, but that studies have not considered the processes through which evaluative experiences contribute to writing. The design of this study is underpinned by a theory of action which synthesises formative assessment with SRL in the context of academic writing. In this chapter, I will introduce the theory of action, the rationale and the research questions. As academic writing is both the process and product of disciplinary learning, SRL is used rather than a cognitive model of writing. Research using models such as Hayes and Flower (1980), Hayes, (2012) and Galbraith (2009) tends to look at the micro-process level of writing skill, rather than how writers manage processes and challenges. SRL is more apt for examining writers' intentions and decisions in response to the particular challenges of academic writing.

The chapter opens by introducing the theory of action, and grounds the model in key works of influence: Sadler (1989, 2009a, 2010), Winne and Hadwin (1998, 2008) and Galbaith (2009). I then introduce the research questions and their theoretical rationale. Next, I present concepts significant to the theory of action and research questions: conditions (including self-efficacy beliefs), task perceptions, students' own metacognitive standards, and large-scale adaptation – changes and adaptation made between writing tasks.

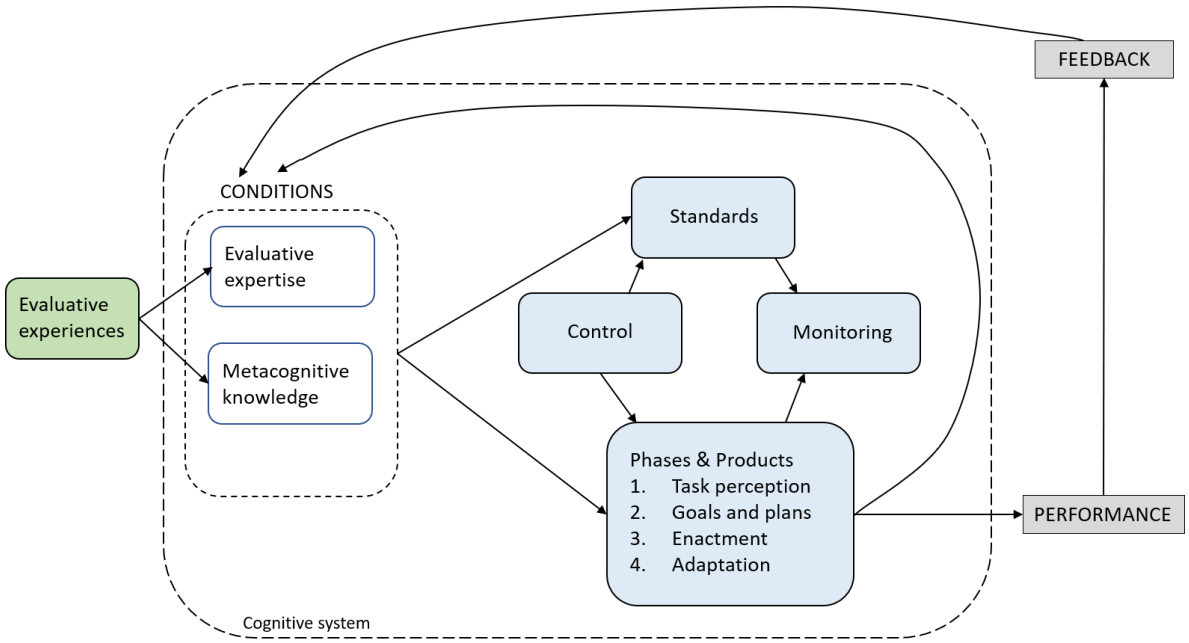
3.2 Theory of action

The theory of action, summarised on the next page in Figure 1, synthesises Winne and Hadwin's (1998, 2008) model of study as self-regulated learning with key concepts from Sadler (1989) and (2010). The figure shows the role of evaluative expertise in a writing episode. In common with other models of SRL (e.g. Zimmerman, 2000; Pintrich, 2000), Winne and Hadwin (1998) see study as unfolding over recursive and overlapping phases. The value of their model of SRL for this study is that task perception, also highly important to cognitive models of writing, is conceptualised as a first stage, distinct from goals and planning. According to Hadwin and Winne (2012, p.204), this is in order to 'highlight two different studying problems learners may encounter: (1) not understanding the academic task, and (2) not knowing how to approach the task they perceive'. These studying problems are common in the academic writing context: task descriptions are often intentionally open and fuzzy, there may be differences between students' and tutors' perceptions of what the task involves. From Sadler, the theory of action takes evaluative expertise (1989) and the inter-related concepts of task compliance, quality and criteria (2010). These are discussed further in the next section; I now outline the theory of action.

Reading the figure from left to right, we begin with evaluative experiences such as peer assessment (in green). Through reading and evaluative a range of work, the student builds evaluative expertise, or richer ideas about quality and salient criteria. Reading others' work may also challenge or develop the student's metacognitive knowledge about writing. When the student comes to produce her/his own essay, evaluative expertise and metacognitive knowledge form part of the conditions of writing (dotted box). The conditions of writing influence how self-regulation unfolds in two key ways.

Figure 1: Theory of Action

Evaluative experiences such as peer assessment build writers' evaluative expertise, which in turn influences task perceptions and the standards students use to monitor and control writing.



First, the conditions of writing shape the phases of writing (in blue). The first phase of a writing task is task perception. The way the student interprets the task is shaped by her/his metacognitive knowledge and evaluative expertise, and in turn determines the plans and goals the student sets, and her/his enactment of writing. With developing evaluative expertise, task perceptions become more congruent with what teachers expect and value.

Second, evaluative expertise and metacognitive knowledge also feed into the self-set standards students hold and apply for metacognitive monitoring and control, which become richer and more detailed. As the basis of metacognitive monitoring throughout the writing episode, standards shape the decisions the student makes. For example, the student may decide to change approach while writing if s/he isn't meeting her/his standards, or s/he may decide to readjust or abandon certain standards if they are too ambitious. Depending on her/his evaluations of writing throughout and at the end of the task, s/he may decide to make adaptations to her/his approach

in future tasks.

As the student writes and come to complete the task, s/he continues to add to and refine her/his metacognitive knowledge about writing. This knowledge, along with evaluative expertise, frames how s/he both reads and appraises work by others and relates to her/his own writing in the future.

Once the writing is complete, the writer may receive feedback on performance, or the finished written product (in grey). This feedback becomes part of the conditions that shape how students self-regulate future writing episodes. Feedback itself is filtered through students' own experience of the task, evaluative expertise and metacognitive knowledge (Butler and Winne, 1995).

3.3 Background to the theory of action

The theory of action provides an explanation for what Sadler argues is the result of offering students evaluative experiences: 'intelligent' self-monitoring (1989, p.143). Sadler defines this as being able to 'monitor the quality of their work during its actual production... making conscious judgements on their own, without help from teacher or peers' (Sadler, 2009a, p.48). Although Sadler does not cite the metacognition and SRL literatures, there are clear resonances here: the aim he describes is productively self-regulated learning. In the writing literatures, too, self-monitoring is an important component of writing, for example the revisions made during writing (Hayes, 2004). It is worth noting that Sadler's definition of self-monitoring is different to a metacognitive definition of monitoring. Like his definition of feedback, 'it is assumed that there has to be some closure of the gap for… self-monitoring to be labelled as such' (Sadler, 1989, p.122). In other words, students have to notice a gap, and act, and so his definition of monitoring therefore subsumes monitoring *and* control, in metacognitive terms.

In order to self-monitor their work, students need evaluative expertise. Sadler does not explicitly define evaluative expertise, so the study draws on my own working definition, derived from Sadler's work: rich knowledge about quality and the salience of criteria which enables students to make qualitative judgements on work. Sadler uses this term in the 1989 paper and 2009 chapter, but not in the 2010 paper. In the 1989 paper, Sadler also uses the term 'evaluative knowledge'. Again, no explicit definition is given, but it is positioned as aligning with teacher evaluative knowledge and enabling students to make complex judgements:

> '…providing guided but direct and authentic evaluative experience for students enables them to develop their evaluative knowledge, thereby bringing them within the guild of people who are able to determine quality using multiple criteria.' (Sadler, 1989, p.134).

In the 2009 chapter, evaluative expertise is again not defined, but it is linked to connoisseurship, which is given as 'highly developed form of competence in qualitative appraisal' (Sadler, 2009a, p.57). In the 2010 paper, neither term is used, but Sadler highlights three concepts which are prerequisites to monitoring. The first, task compliance, is defined as: 'the congruence between the type of response stipulated in the task specifications and the type of response actually submitted by a student' (2010, p.543). Developing metacognitive knowledge about what makes work meet requirements is essential for learning to be rewarded by assessment (Butler and Cartier, 2004). For complex academic writing, there are many facets that make up task compliance. Failing to answer the question is a perennial task compliance issue, and one that crops up frequently in Examiners' Reports at Oxford (e.g. Faculty of English Language and Literature, 2013 – hereafter ELL). Genre knowledge is also important. For instance, English Literature degree exam essays, tutorial essays and A-level essays share some characteristics and moves, but differ enough that producing one genre where another is required would mean the task requirements are not fulfilled.

The second and third concepts, quality and criteria, are likely to be tacit and discipline-specific. Quality and criteria are often conflated: criteria are seen to stand for quality. However, Sadler (2010, 544) sees quality as a unitary construct: 'the degree to which a work comes together as a whole to achieve its intended purpose.' He argues that:

> 'In practice, quality is often easier to recognise when it presents itself than it is to define in the abstract, or account for fully in the particular. Not uncommonly, something significant is lost when attempts are made to express quality in propositional or declarative form, that is, in words, including rubrics and expansions of fixed criteria.' (Sadler, 2010, p.544)

Recognising quality precedes criteria because students need direct experience of works in order to appreciate the different criteria at play, and when particular criteria are salient and when they are not. Rubrics cannot substitute for this knowledge, as discussed in Section 2.4.2. According to the theory of action, evaluative experiences should help students to develop the capability to recognise particular criteria in their own writing as well as the writing of others, and to develop a metalanguage. Thus, peer assessment is a kind of experiential learning that allows students to build the tacit knowledge about writing that contributes to richer task definitions and standards that are better aligned with what is valued in the discourse community (Swales, 1990).

The other basis for the theory of action, Winne and Hadwin's (1998, 2008) COPES model of SRL, consists of four stages which are recursive and overlap. Students' standards in relation to Sadler's three concepts of task compliance, quality and criteria can be drawn on throughout writing. The phases are not rigid phases; the last in particular may be missed out. In the first stage, the student defines (1998), perceives (2008) or represents the task (Hadwin and Winne, 2012). Secondly, the student sets up goals and plans. Thirdly, the student enacts study, drawing on tactics, and finally they may make adaptations to their future approaches to studying. COPES is an acronym for five key facets of SRL that occur across the stages. Throughout learning, students respond to

conditions, both internal (e.g. motivation and beliefs) and external (e.g. time constraints), enact *operations* which create cognitive, affective and motivational *products*, which are then *evaluated* using *standards* or criteria (Hadwin and Winne, 2012). Beliefs about task compliance, quality and criteria are some of the conditions of learning the student brings to the task, and inform the standards students use to judge writing as it unfolds. To give the example of writing an essay, on receiving the task, before getting into the writing proper, the student may produce an interpretation of the task or question. This task perception phase is shaped by task conditions, e.g. time constraints, as well as metacognitive task knowledge and beliefs about quality, for example. The student's task perception is produced through carrying out operations such as reading the instructions, or annotating the question for ideas. The student may evaluate their task perception using standards originating from previous experiences of writing, their own beliefs about quality and which criteria are salient (Winne and Hadwin, 1998, p.280). For example, the student may re-read the question in order to check their understanding of the task. If the task is very routine or straightforward, the task perceptions stage may be fleeting or even skipped. However, in academic writing, developing task perceptions may take some time, and this phase may also recur throughout writing, as evaluations made while self-monitoring may indicate that the original interpretation has changed, or isn't being followed through. The writer then may make plans or goals for the writing, for example, outlining a paragraph structure or deciding which material to include. Enactment is then the main phase of writing. In this study, it is counted as the first full sentence of the writing task, a clear and visible demarcation, although the student may revisit plans and goals during writing. As they write, the student will continue to self-monitor. With each evaluation against standards, there is an opportunity to self-regulate, if necessary: by altering conditions, operations or standards (Winne and Hadwin, 2008, p.306). For example, the writer may realise on finishing a paragraph that they haven't met the standard of quoting from a primary text,

and rather than continuing with the operation of writing the next paragraph, decide to re-read before continuing. Alternatively, they may decide to alter or abandon their standard that a quotation is required in every paragraph, and continue writing. Such evaluations are not just related to cognitive products: ‘as students move through the phases of SRL, they are constantly and inherently experiencing affect with every product they create’ (Winne and Hadwin, 2008, p.306). Again, students can self-regulate, for example by using self-talk to keep them going through a task – an operation – or adjusting standards to make the task more feasible. Once finished, evaluating how the writing went may lead the student to decide to make adaptations to her/his approach in future tasks: large-scale adaptation. The writer may also receive feedback on their performance, which together with the experience of writing the text, feeds into the conditions of the next writing episode.

The phases of the COPES model are similar to those in other SRL models (Zimmerman, 2000; Pintrich, 2000). However, the value of this particular model of SRL is that task perception is presented as a separate stage to goals and planning. Also, as Greene and Azevedo (2007, p.338) point out, 'with monitoring and control functioning as the hubs of regulation within each phase, Winne and Hadwin’s model can more effectively describe how changes in one phase can lead to changes in other phases over the course of learning'. This emphasis on the importance of metacognitive monitoring and control, aligns with the governing definition of writing used in the study, Hacker et al.'s (2009) conceptualisation of writing as applied metacognition.

The theory of action goes beyond the Winne and Hadwin model (1998, 2008), originally designed to describe self-regulated learning in the context of studying, to consider the particular challenges of academic writing. In the focus on task perceptions and standards, the SRL model aligns well with cognitive models of writing such as Hayes and Flower (1980) and Bereiter and Scardamalia (1987).

While such models focus on the micro-processes that occur as we compose, SRL encapsulates the fine-grained, but also provides an overview of how students can strategically manage writing. A further limitation of these models of writing are that they do not consider how ideas are discovered during the writing process; rather, they treat ideas as retrieved from long-term memory and translated into text: thinking and production are considered to be separate processes. Galbraith's (2009) dual process model offers an alternative view. He argues that there are two systems at play: intentional, goal-directed knowledge-transforming, similar to that described by Bereiter and Scardamalia (1987), and knowledge-constituting. The latter system is not goal-directed; rather ideas are generated through implicit connections and syntheses made through writing itself. Knowledge-constituting leads to discovery and accounts for the phenomenon of writers learning through writing. It is likely to be at play during academic writing as tasks require the handling of complex knowledge. Galbraith (2009) argues that good writing is likely to draw on both processes. Knowledge-constituting can be harnessed strategically for idea generation by using strategies such as free writing (Elbow, 1998), that is, writing without pause, goal or filter for ten minutes.

The two systems present a challenge, arguably more prominent than that of working memory, that of 'the two conflicting impulses: the impulse to maintain a pre-determined plan, and the impulse to follow an unfolding train of thought' (Galbraith, 2009, p.24). In a time-limited essay such as an assessment situation, the two systems are likely to present multiple self-regulatory challenges. What to do when a new idea bubbles up in the middle of an essay? Is it best to follow this idea, or stick with the plan? Such decisions are likely to be made according to writers' standards to do with, for example, the value of the idea, the coherence of the essay, and beliefs about whether following the plan will lead to a better outcome. To manage this, and other challenges, such as working memory (Kellogg, 1994), writers may plan heavily before writing, crystallising their ideas, or set

aside time to revise intensively (Galbraith and Torrance, 2004). In terms of the Winne and Hadwin (1998, 2008) model, some revisers may skip the plans stage altogether and allow knowledge-constituting to take precedence for a time. These two overall strategies may represent students' usual approaches to writing (Torrance, Thomas and Robinson, 1999; Galbraith and Torrance, 2004). Tasks that are adapted to match the 'planner' or 'reviser's' overall strategy will support learning and writing quality more effectively than tasks that force writers to take a different approach (Kieft, Rijlaarsdam, Galbraith and van den Bergh, 2008; Kieft, Rijlaarsdam and van den Bergh, 2008). In the context of this study, the timed conditions of examined writing may interfere with writers' usual preferences, as they may not feel they have the latitude to plan or revise as thoroughly as they might like. Other goals or standards, like demonstrating knowledge, or writing as much as possible, might take precedence.

The goals set during study and those set for writing are likely to be quite different. A frequently used definition of goals in the SRL literature is 'the object or aim of an action, for example, to attain a specific standard of proficiency, usually within a specified time limit' (Locke and Latham, 2002, p.705). Study may involve goals such as memorise and recall a list of vocabulary. Such goals are perhaps more clear-cut and easier to articulate than writing goals, particularly in the timed writing examined in this study. Flower and Hayes' (1980 p.27) seminal comparisons of expert and novice writers' think-aloud protocols of writing uses the synonyms 'purposes' and 'intentions' for goals. They found that writers set up four types of goals: the effect on the reader, the writerly persona or voice, creating meaning and genre. In an hour's timed essay, such goals are likely to be set up and met quickly: they are micro-goals in comparison to study goals. Such writing goals are also likely to be similar between individuals: to a large extent, the over-arching communicative goals of examination writing are fixed by the assessment situation and genre: students aim to write a 'good' essay to convince examiners of their knowledge and their argument. Of course, holding

such goals requires understanding of the assessment situation and genre, and students can still set goals within this, for example to cite a favourite critic, or use a particularly good point, or try out a new argument. But all students would presumably share the goal to finish the essay within the time limit. The standard for what constitutes a finished essay would differ from student to student, though, and can be adjusted as students write.

An additional consideration is that the term 'goal' is not part of the metalanguage used in school, and to an extent, university writing. The national curriculum presents writing as an idealised, linear process of planning-writing-revision, and children are taught to plan before they start writing (Myhill and Jones, 2007). So, writers' intentions are more likely to be expressed through 'planning', which may function both as an expression of essay goals – material and arguments to include, for example – and as a strategy to organise the writing process, as discussed above. Thus, as discussed in the following sections, standards – beliefs or conceptions of good writing that are actually used to judge and monitor learning – are perhaps a better lens.

3.4 Research questions

The theory of action sharpens the broad aim of this study, which is to investigate what and how students can learn about writing through appraising examples, writing a practice essay, reviewing peer-written essays and receiving peer feedback. The over-arching research question (RQ) is: (how) does students' self-regulation of writing develop in the context of peer assessment writing workshops? The more specific research questions used to answer this are as follows:

RQ1. During the workshops, which criteria do students use to evaluate written essay products?

RQ2. Which criteria do students use as standards to self-regulate their essay writing processes?

RQ3. (How) do students' task perceptions shift over the course of the study?

RQ4. (How) does peer feedback contribute to students' self-regulation of writing?

RQ5. What is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing?

RQ1 taps into students' evaluative expertise by asking which patterns can be discerned in student writing, identifying underlying criteria. As criteria were not pre-specified in the workshops, it is important to identify which constructs underlie students comments about writing – their own and others. As Sadler (1989) points out, there can be an infinite number of criteria for writing, so identifying the key criteria that students use is important for understanding students' evaluative expertise. The convergence of these criteria with official criteria will also be discussed, addressing whether as a group students miss any criteria that are valued by examiners. RQ2 examines whether and how students use criteria to self-regulate their learning: which criteria become (meta)cognitive standards for writing. It can be hypothesised that not all criteria will be useful for writing, and that it may not be feasible for students to adopt all the moves and strategies that they recognise as desirable into their own standards.

RQ3 asks (how) do students' task perceptions shift over the course of the study? As discussed in the next section, task perceptions are particularly crucial to self-regulated writing as they influence

how the process unfolds. In English literature, very little about essay questions may be explicit, and task perceptions will influence the extent to which an essay achieves task compliance. RQ4 moves to post-writing, focussing on feedback. In this study, feedback is conceptualised as one aspect of the formative assessment intervention, rather than the dominant aspect. It is one event contributing to the conditions that shape future self-regulated writing. Students may or may not take up what is suggested in feedback, and their receptiveness to feedback may be shaped by whether the feedback taps into shared standards about writing. So, in order to recognise that students have and do exercise choice in how feedback is used, RQ4 is phrased as (how) does peer feedback contribute to students' self-regulation of writing?

The final research question acknowledges that self-regulation is not just a metacognitive process: it involves powerful beliefs and affect. Self-efficacy beliefs shape and are shaped by writing experiences. RQ5 asks: what is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing? The peer assessment workshops may offer several sources of self-efficacy, as discussed later in the chapter. The research question thus tests the hypothesis that students' self-reports of self-efficacy will change between the first and final workshop.

Together, these research questions provide a multi-faceted response to the over-arching research question: (how) does students' self-regulation of writing develop in the context of peer assessment writing workshops? There is not, however, a research question to address writing quality. Writing development is not linear, and trying out new strategies does not automatically lead to an improvement in writing quality, as judged in the final product (Camp, 2012). Indeed, learning cannot always be judged from performance (Soderstrom and Bjork, 2015). Researchers have called for strategy use to be taken into account when examining writing development (Alamargot and

Fayol, 2009; Slomp, 2012). Thus, the emphasis on how aspects of self-regulated writing develop is more meaningful for this study. The methodological difficulties of assessing writing quality in the particular study context is discussed in the next chapter (section 4.11.7). The discussion now addresses aspects of the theory of action important to the research questions.

3.5 Conditions

Conditions consist of both task conditions such as time, and what the learner brings to the task in terms of his/her beliefs, metacognitive knowledge, topic knowledge and any other baggage (Winne and Hadwin, 1998). Flavell (1979) originally divided metacognitive knowledge into three types: person, task and strategy. These types still stand today (Harris et al., 2009). Metacognitive knowledge 'encompasses everything that you could come to believe about the nature of yourself and other people as cognitive processors.' (Flavell, 1979, p.907). In the context of this study, this can include knowledge about your strengths and weaknesses as a writer, or how long it generally takes to handwrite a page of essay in an exam situation. Such metacognitive knowledge feeds into self-efficacy beliefs, which are prospective – how well you believe you can perform a task (Rijlaarsdam et al., 2008). Metacognitive task knowledge is given as an example in the theory of action. This includes task demands like how much revision you need to do to write an essay. Metacognitive strategy knowledge involves the types of strategies most likely to be helpful in achieving goals. For example, it may include the knowledge that outlining the argument at the beginning of an essay is effective. Or, it may be founded in experience, such as finding mindmapping literary quotations more effective and enjoyable than trying to memorise them using index cards. Metacognitive knowledge is updated by metacognitive experiences, although we may not always remember or make good judgements from such experiences (Flavell, 1979).

Sadler's (1989) notion of evaluative expertise is also included as a condition – judging quality, and knowing which criteria are salient in the particular context. Although the line between cognitive and metacognitive is often blurred, evaluative expertise when judging others' work is essentially cognitive rather than metacognitive knowledge. For example, noticing that the writer has effectively structured their introduction by including a thesis statement is not metacognitive because it doesn't involve one's own cognition, but rather a written product. However, evaluative expertise can inform students' metacognitive knowledge about their writing. Knowing that others use thesis statements to structure their introductions and that this strategy is successful may influence your appraisal of your own writing, and lead you to reflect on when you have and haven't used thesis statements in the past. It may lead to you setting 'use thesis statement' as a standard in the next essay.

It is important to highlight while the theories drawn on here privilege the (meta)cognitive, emotional experiences are important to self-regulating writing. Flavell's (1979) initial definition of metacognition included affective experiences. Winne and Hadwin's model recognises this: 'as students move through the phases of SRL, they are constantly and inherently experiencing affect with every product they create' (2008, p.306). Such (meta)cognitive feelings, although transient and fleeting, influence metacognitive monitoring and control, whereas emotions such as interest or anxiety either motivate engagement more broadly, or present barriers (Efklides and Volet, 2005). Research on self-regulation does not often combine the metacognitive and emotional. Boekaerts' (2011) model is an exception, foregrounding emotion and emphasising how students' purposes during learning are not always focussed on the task but rather on the self, and avoiding discomfort or threat. In other words, students may take a growth pathway, and embrace challenge, or take a self-protective well-being pathway. More recently, it has been shown that self-regulating learners use a range of strategies to regulate emotion as they study (Webster and

Hadwin, 2014). While it is beyond the scope of the study to examine emotion as a key focus, I want to acknowledge the role of emotion in metacognition and writing, and in particular in high-stress writing situations like Prelims examinations. Assessment itself is emotionally charged (Boud and Falchikov, 2007), as the considerable literatures on examination stress and test anxiety show (e.g. Putwain, 2009). A particular issue that arose once data collection commenced was confidence, or self-efficacy beliefs, discussed next.

3.6 Self-efficacy beliefs

In the theory of action, beliefs about writing are both conditions influential during writing and can be updated by the writing experience. As described later (Section 4.7.1), it became apparent early in data collection with students and tutors that 'confidence', or self-efficacy was perceived to be important to writing success in the particular context of the highly pressured Prelims exams. Self-efficacy beliefs are 'beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments' (Bandura, 1997, p.3). Self-efficacy beliefs are domain-specific. The relationship between self-efficacy beliefs and achievement is not necessarily direct; rather self-efficacy can shape the choices writers make. Pedagogical approaches that foster perceived self-efficacy also foster writing achievement (Schunk, 2003). More broadly, a recent meta-analysis found a moderate and positive relationship between academic self-efficacy and achievement in 59 studies (Honicke and Broadbent, 2016). This research also highlighted that the relationship between self-efficacy can be mediated by other factors, including strategy use and goal orientations. In writing, research has shown how self-efficacy shapes students' approaches to writing. Lavelle and colleagues' survey work identified five scales or profiles of writing, including a low self-efficacy approach, which was also associated with surface rather than deep writing

strategies (Lavelle and Zuercher, 2001; Lavelle and Guarino, 2003; Lavelle, 2009). These findings reflect trends across wider research that students with low self-efficacy beliefs are unlikely to self-regulate their learning productively (Zimmerman, 2011). More recent work suggests that self-efficacy for writing is not unidimensional. Survey research with school and university writers found that ideation, self-regulation and conventions were separate factors (Bruning, Dempsey, Kauffman, McKim and Zumbrunn, 2013). So, for example a student may feel they can generate ideas for writing and use correct spelling, punctuation and grammar, but less able to plan and revise, or self-regulate writing confidently.

There are four key sources of self-efficacy: mastery experiences, observing others, social persuasions such as positive feedback or encouragement, and physiological and emotional signs, e.g. interpreting a feeling of dread in the pit of the stomach as a sign that one cannot complete the task (Bandura, 1997; Pajares, 2008). Mastery experiences give students the strongest evidence of their self-efficacy, as found in van Dinther, Dochy and Segers' (2011) review of a broad range of interventions across higher education, and Pajares, Johnson and Usher's (2007) writing-specific survey research with school pupils. However, a writing experience is not automatically a mastery experience. In order to interpret it as such, a task has to be appropriately framed, structured so there is an appropriate level of challenge, and students need perceive they are coping and progressing (van Dinther et al., 2011; Bruning and Kauffman, 2015). Where students complete tasks without such framing, they may have to alter their standards and intentionally self-monitor progress so they notice success (Kitsantas and Zimmerman, 2006). The second source of self-efficacy is vicarious: observing others' performances, either through a model or through exemplar work, as discussed in Section 2.7.3 of the literature review. The third, social persuasion, is considered a weaker source of self-efficacy, and indeed the research on praise in feedback is mixed (e.g. Shute, 2008). Educational experiences that build self-efficacy beliefs may offer multiple

sources (van Dinther et al., 2011). As noted in the literature review, the extant research on self-efficacy in formative assessment has been limited, and the inclusion of this research question will enable the study to contribute to an under-researched area.

3.7 Task perceptions

Task definition, representation or perception is an important process in both the SRL and writing literature. In their seminal conceptualisation of writing as problem-solving, Flower and Hayes (1980, p.23) stress that 'writers themselves create the problem they solve' and that setting up the problem 'may be one of the most critical steps the average writer takes'. Task perceptions are constructed in the interaction between the task and its context, and what students bring to the task (Flower et al., 1990; Winne and Hadwin, 1998, 2008). In this interaction, 'students construct a personalised profile of the task at hand' (Winne and Hadwin, 2008, p.298). The cognitive conditions students bring to the task includes subject knowledge, metacognitive genre and task knowledge but also students' more general conceptions about the nature of learning and of the task. Butler and Cartier (2004, p.1740) consider that 'the relationship between conceptions and metacognition is reciprocal because although more generalized task conceptions are grounded in metacognitive knowledge, conceptions in turn influence subsequent re(constructions) of metacognitive knowledge through their impact on self-regulation'.

The task perception process necessarily starts at the beginning of the writing episode, but is characterised by both Winne and Hadwin (1998) and Flower et al. (1990) as evolving as the task unfolds. Flower et al. (1990) see this as process of metacognitive noticing and evoking from long-term memory. There is empirical support for the evolution of task definition over the course of a study episode from Greene et al.'s (2012) investigation of science learning in a hypermedia

learning environment, and from (Nelson, 1990) and Negretti (2012) for shifts in students' writing task perceptions over the course of a semester. However, in short, time pressured tasks, clarifying task perception early on appears to be important. For instance, Van den Bergh and Rijlaarsdam's (2001) study of the temporal distribution of writing processes during a relatively short task found that task perception mostly occurred in the beginning phases of writing. Task perception correlated positively with text quality when it occurred at the beginning of the process, but not at the end. In Flower et al.'s (1990) study, dramatic shifts in task perceptions were also problematic, particularly if the writer did not notice.

Hadwin and Winne (2012) theorise that task perceptions operate on three different layers: the explicit, implicit and socio-cultural. As this section will show by giving example items from A-level and university examinations, the implicit and socio-cultural layers in particular come into play in academic writing about English literature. Explicit features of tasks are those which are fairly straightforward to define and are often discussed in classroom settings (Hadwin and Winne, 2012). For example, in the AQA A-level English Literature B Unit 3 examination in June 2013, students were instructed to 'answer two questions. Answer one question from Section A and one question from Section B.' In contrast, implicit task features have to be 'extrapolated from external and internal conditions and include task purpose, connections between task and other course components, relevant task resources, and inferences about kinds of thinking to be demonstrated' (Hadwin and Winne, 2012, p.206). Literature examinations commonly provide quotations alongside each question. In the following A-level item, the candidate is directed to the quotation, and would need to infer that they need to engage with both 'evil' and 'ambition'. Even though these terms are in everyday usage, candidates would also need to know to use genre specific vocabulary in their own answer.

> "Macbeth is a play about the nature of evil rather than the nature of ambition."
>
> To what extent do you agree with this view of the play? (AQA A-level English Literature Specification B, Unit 3, June 2013)

By the time they sit Oxford English Finals, forming task perceptions is more demanding. Students are presented with quotations from authors or critics, but offered no explicit guidance for how to engage with them.

> IAGO: Do it not with poison, strangle her in her bed – even the bed she hath contaminated.
>
> OTHELLO: Good, good, the justice of it pleases; very good!
>
> (Othello)

And:

> 'The multiple marriages that conventionally end comedies operate as symbols for full participation in life' (SUSAN SNYDER).
>
> (Honour School of English Language and Literature Course I, Paper 2, Trinity term 2013).

These questions may be unintelligible to disciplinary outsiders: apart from requiring rich literary knowledge, they require task interpretation at Hadwin and Winne's (2012) socio-cultural level. Such features 'are sampled from the broader realm of courses, program and discipline. They are embedded in the discipline and in the academic program. These features extend beyond the specific task and course, drawing on disciplinary genres and dominant disciplinary epistemological and pedagogical beliefs' (Hadwin and Winne, 2012, p.206). In the first year Prelims examinations, some further guidance is given in that themes are identified for the students. However, what forms of 'writing on...' are appropriate and rewarded is embedded in the disciplinary context:

> '[How] little the habit of writing in the serial form encourages one to read in that odious way, which so many simple folk, thank heaven, think the best' (HENRY JAMES)
>
> Write on the dissemination and reception of texts AND/OR questions of class in any writing of the period.
>
> (Preliminary examination in English Language and Literature: Paper 3. Literature in English (1830 -1910), 2013).

Thus, the exam essay is a reading-to-write task (Flower et al., 1990). Not only do students have to transform, shape, construct the knowledge from the reading they have done throughout the year, but they also have to respond to the short text that forms the question: the quotation. The examiners' reports indicate that diversity in interpretations of the quotation is valued – candidates are praised for going beyond the standard authors and more obvious approaches (ELL, 2013). As well as interpreting the quotation, deciding how it is to be used in the written response is also key. Flower et al. (1990) term this the 'organising plan' for the writing. Some possible ways of representing and using the quotation in responses are:

- As a springboard for own thoughts about a theme or text
- Identifying key words and responding to these
- Mentioning it in the introduction and coming back to it in the conclusion
- Connecting words or imagery in the quotation with another text
- Using it to frame an argument
- Contextualising the quotation within the text / historical period / critical tradition it came from
- Using the quotation as a starting point for a synthesis on a key idea
- Ignoring the quotation and responding instead to the wording of the rubric / question

In these ways, the quotation is used – or not – to shape, or frame or even transform the literary knowledge the student has accumulated over the course of the year. Not all of these ways of

representing the quotation and the task would be valued by examiners. For instance, using the quotation as a springboard for thoughts about a topic or author could lead students into a knowledge telling strategy where they recall important points but don't necessarily consider the audience's need for these to be shaped into a coherent argument (Bereiter and Scardamalia, 1987). From the perspective of a student who represented the task in this way, this approach could represent a valid and effortful response to the quotation, particularly if the choice of items on the paper is not well matched to their knowledge. However, a teacher or examiner may not agree.

The insights on task perceptions discussed here from the SRL and cognitive writing literatures resonate with findings from the academic literacies research, for instance Lea and Street's (1998) conclusion that the in/appropriateness of student writing often has less to do with explicit, generic writing skills, for instance, punctuation, and more to do with subject epistemology and the academic frameworks in which lecturers are experienced. Leaving the implicit and socio-cultural aspects of tasks tacit means that some students who are not yet experienced in these aspects are disadvantaged, as Myhill (2005) found in her analysis of Year 6 SATS tasks. According to the theory of action, appraising authentic examples of work through activities such as peer assessment should help students develop their metacognitive task knowledge and thus develop task perceptions that address more than just the explicit, so making it more likely they will answer the question in a way that teachers or examiners might expect, or achieve task compliance (Sadler, 2010).

3.8 Standards

Standards are the criteria students use to monitor and evaluate their learning as it unfolds (Winne and Hadwin, 1998; 2008). It is important to disambiguate the term ‘standards’ as it is in wide yet wildly differing use in educational research. Most broadly, ‘educational standards’ refer to ‘the quality of educational provision’, and the standards of public examinations are the subject of intense scrutiny and debate as they are used as a proxy to judge the success of the wider system (Cresswell, 2000, p.69). The higher education policy context, too, has shifted towards learning outcomes and threshold standards, and these are being put to similar comparative and accountability uses at programme and institution levels (Boud, 2017). Within courses, an ‘academic achievement standard’ can be defined as a ‘minimum achievement level used as a reference point when judging the quality of a student’s work’; it is distinct from a criterion, which is simply ‘a property or quality’ (Sadler, 2014, p.275). Academic achievement standards are set by communities of examiners or teachers, and communicated publicly. In contrast, the standards that students use to metacognitively judge and monitor their own work are personal, idiosyncratic and often tacit. While achievement standards are used by assessors to judge the quality of a performance, ‘multifaced’ metacognitive standards are are generated and applied during each phase of the process of learning (Greene and Azevedo, 2007, p.336). Poitras and Lajoie (2013, p.225) define standards as 'the learner's beliefs regarding the requirements that have to be met in order for learning to be deemed a success'. So, metacognitive standards may be related to achievement standards, but they cannot be taken as identical – any external standard will be filtered through the learner’s interpretations and beliefs.

In addition, Winne (2014) emphasises that standards apply not only to the products of learning, for example a plan or essay, but other aspects of COPES. Students can even modify their standards for

standards, for example deciding that one standard is too ambitious, or that the way they are judging learning is not working out. So, standards are contingent and can shift during a task. It is important to disambiguate standards from goals. Taken together, a set of standards makes up a goal (Winne, 2014). In this study, the emphasis is on standards rather than goals as in an exam essay situation, students have less scope to set goals, and goals may be more tacit than in entirely self-directed tasks. Standards are thus more interesting for investigating individuals' proactive decisions, and there is more likely to be intra-individual variation.

As standards are not necessarily fixed at the beginning of a task, the products of phases such as task perceptions can be then in turn used as standards (Winne and Hadwin, 2008). For example, if a student develops a task perception that the question is about showing the importance of a key word in literature, this task perception might become a standard, and s/he may monitor the use of the key word throughout writing the essay. However, monitoring is more likely to lead to success in a task when standards align with what is valued in the context. Simply responding to and using a key word throughout the essay may not be judged to be sufficient; multiple standards may emerge from different layers of task perceptions. For example, the standard for using a key word may be complemented with a standard for creating a thesis or argument from this key word, and these may be used in combination.

There is little research into the nature of students' standards. The section in Greene and Azevedo's (2007) theoretical review of Winne and Hadwin's model is noticeably thinner than other aspects, but since its publication there have been advances. In the field of metacognition, research in metacognitive judgements and monitoring has traditionally involved experiments where participants study information then take a test. Schraw (2009) gives a taxonomy of measures used in such research to tap into students metacognitive monitoring, which include Judgments of

Learning, judgements that participants make about how well they will recall information that they have just studied, and Online Confidence Judgments, which are made during testing. However, as Hadwin and Webster (2013) point out, real world learning with ill-defined tasks will involve more complex judgements, and students often have to rely on self-set rather than external standards. Much research in this area focusses on calibration, which 'reflects the degree to which individuals' judgments about their understanding, capability, competence, or preparedness correspond to the understanding, capability, competence, or preparedness they actually manifest' (Alexander, 2013, p.1). Calibration is important to metacognition because students' judgements influence how they self-regulate. If judgements are out of kilter, then a learner may make poor decisions. In general, beginning students are thought to be too generous in their metacognitive judgements, although those with more expertise are thought to judge overly harshly (Dunning, Heath and Suls, 2002). Such patterns are attributed to lack of metacognitive skill, but what underpins poor metacognitive judgements is not ability, but a thin set of standards. How can a novice who has never seen a full range of quality of work or experienced expertise first hand possibly know that their work is not very good? The self-assessment literature reflects these trends (Boud and Falchikov, 1989; Dochy, Segers and Sluijsmans, 1999). More recent research has shown that student self-assessments can become more closely aligned with teacher assessment, even as quickly as over one semester, although students who over-estimate their achievement are less likely to become more calibrated (Boud, Lawson and Thompson, 2013).

Metacognition and SRL research is only just beginning to address how students make judgements – which sources of information they draw on. Such research is important, as according to Dunlosky and Thiede (2013, p.59) 'just because a student is asked to judge a particular aspect of his or her own cognition, does not mean the she or he has any particular insight into that aspect of cognition.' Thus research is needed to identify the basis for judgements, and to identify both

productive and problematic sources of metacognitive monitoring. Dinsmore and Parkinson (2013) asked undergraduate students in a research methods class to read a textbook passage, then respond to 10 items requiring recall. They marked confidence ratings on a 100mm line to convey how successfully they had answered these items, and wrote an open-ended response as justification for two items. They repeated the exercise with a different passage a week later. Coding of the open-ended responses showed that students' confidence ratings were based on their prior knowledge of the material, the texts, the characteristics of the item, and guessing. Students sometimes drew on multiple sources to make their ratings, but exploratory analysis indicated that this did not necessarily lead to greater calibration accuracy.

Students may be set external standards, for example through assessment criteria. Few studies have explicitly studied how standards shape students' self-regulated learning. Kitsantas and Zimmerman (2006) conducted an experiment on the acquisition of dart-throwing skill with university students, comparing three conditions, as well as the effect of graphing. The absolute standard group were given points scores for bull's eye versus any other hit. The graduated standards group were told point scores for each concentric circle. The control group were told that the aim of the exercise was to hit the bull's eye. In addition, there were two sub-conditions for the experimental groups: graphing and no-graphing. Those in the graphing groups recorded their scores at the end of practice; others did not. At post-test, participants who were set graduated self-evaluative standards and also graphed significantly out-performed others in terms of dart skill and motivational beliefs, including self-efficacy. Kitsantas and Zimmerman (2006 p.211) consider that 'often students set unrealistic standards for themselves, and instructors need to identify these students and help them readjust their standards to be sensitive to small improvements in skill.' They argue that graphing progress on graduated standards made progress more visible, motivating students to make changes where necessary, and framing learning more positively.

So far, only one calibration study has been conducted with extended writing. Hawthorne, Bol and Pribesh (2017) looked at calibration accuracy for students' predictions and postdictions of writing performance on a writing competence exam that was a prerequisite for batchelor's level graduation from a university in the US. Little is reported about the task, simply that it was argument-based, with three hours to write 500-600 words. Students were asked to predict the score out of 20 they would receive on the essay before receiving the task and after completing it. The 596 participants were assigned to three conditions. In the global condition, students were given a scale from 1-20, where 1 was labelled 'incompetent' and 20 'superior competency'. In the global and criteria condition, students were given the same global scale, and were also asked to predict/postdict the score they would achieve on four criteria. These were presented on a scale from 1-5, with each labelled, again from 'incompetent' to 'superior competency'. Participants in the rubric condition were again asked to predict their score for both global and criteria, but were also given a detailed rubric which gave a detailed explanation of the performance required for each of the four criteria on scale of 1-5. The authors measured calibration accuracy as the participants' predictions and postdictions of scores, as compared with the scores allocated by trained raters. Unsurprisingly, there was very little and no significant variation between the conditions in calibration accuracy. On average, students' predictions and postdictions over-estimated performance by 4 points.

From an assessment perspective, this study is problematic. There is little reflexivity around the tools (or lack of) that students were asked to use to predict their performance. Moreover, as is usual for writing studies, for the expert raters providing scores, intraclass correlations were not high. For example, agreement on sentence structure was .66, and organisation was .69. There was more agreement on global quality, with an ICC of .81. If raters who have been trained to use a rubric cannot completely agree on quality, then why should we expect calibration accuracy from

students who are asked to come up with scores without training, or even access to any rubric at all? As teachers and other professionals learning to rate writing have to go through extensive training, it does not seem meaningful to measure students' calibrations when they do not have similar evaluative expertise. It is therefore not surprising that there was no difference in calibration accuracy between the conditions.

There is a real research gap, then, in identifying which standards students use to judge their learning and self-regulate, and the origins of such standards. Clearly, the field is not in a position where metacognitive judgements of writing can be meaningfully measured. Thus, the exploratory nature of this study should contribute to the theory and literature. To guide analysis, this study uses a working definition to operationalise standards. It is synthesised from Winne and Hadwin (1998, 2008), Greene and Azevedo (2007) and Poitras and LaJoie (2013) and reads as follows:

> A criterion or belief that is used to judge or evaluate learning, either how it is going in the moment, or how it went (for example, in the self-evaluation part of the interview).

This definition is phrased in order to distinguish standards from general beliefs about writing. A belief only becomes a standard when it is used to monitor or judge learning. Where writers are learning to write in a new genre, they may hold beliefs that are effectively ideals about writing. These may be as yet unattainable, or just not feasible to use due to other constraints. A belief is also only considered a standard where it is used in reference to a particular piece of writing or learning, not writing more generally. This distinction serves to make the often blurred lines between cognition and metacognition a little clearer.

3.9 Large-scale adaptation

A key characteristic of self-regulation is adaptation that is based in metacognitive monitoring. During writing, students may make adjustments to their approach, but in the final phase of writing, they consider future tasks. In other words, 'students deliberate about how to change their methods for completing tasks like this one to close gaps they perceive between the results of their work and goals' (Winne and Hadwin, 2008, p.300). This is analogous to self-assessment in the formative assessment literature, but it is informal and could be fleeting – indeed as Winne and Hadwin (1998) note, it doesn't always happen. Also, the standards which students use to judge the results of their writing are likely to be those used throughout the task – they are self-set, unless students choose to use external standards such as assessment criteria. Classroom-based self-assessment, in contrast, is usually teacher-regulated: initiated by the teacher with criteria or a rubric provided (Vermunt and Verloop, 1999). In addition, while classroom-based self-assessment may focus on the cognitive and written results of a task, from a self-regulated learning perspective students may also adapt and adjust their self-efficacy beliefs, for example, realising that they can do a task well. Also, students may adapt their goals and standards, realising that they were set too low or too high.

Winne and Hadwin (1998, p.285) theorise that 'adaptive decision making reaches beyond the boundaries of the present studying task to change the conditions for studying in the future.' In other words, the evaluations students make about the results of a task carry forward into the future, shaping the conditions for future tasks. For instance, a writer may have tried a new strategy, say, defining her thesis in one sentence before starting to write, and evaluated that this helped her organise the essay, and also feel more self-efficacious about the argument. She might intend to use this strategy in future essays, thus adapting her approach. Such large-scale

adaptation is analogous, then, to Salomon and Perkins's (1989) concept of forward-reaching transfer, where a learner intentionally considers how knowledge and strategies might be applied in a future context. However, transfer isn't sufficient to explain large-scale adaptation: writers may come to more critical conclusions, and decide not to do something again. For example, a writer may have decided partway through the essay that the question they chose did not, in fact, meet their standards for suitability or interest, and as a result, their essay was not going to be as relevant as they wanted. They may reflect that scanning the question paper for keywords was not sufficient, and plan to spend more time thinking through possibilities – developing task perceptions – before making a decision.

Thus, through large-scale adaptation, writers can mitigate against the pitfalls of over-zealous transfer: continuing with old routines which have worked well in new situations, even though they blind the learner to the new (Schwartz, Chase and Bransford, 2012). Although much attention in the university writing literatures has been paid to failure to transfer, for example strategies from first-year composition courses to domain-specific ones (Donahue, 2012), over-zealous transfer is potentially more of a risk in the context of this particular study. As seen in Section 2.7.2 of the literature review, writers can unwittingly transfer genre knowledge into new tasks, even when they think they are writing in a different genre (Bereiter and Scardamalia, 1984; Artemeva and Fox, 2010). In a UK university where there is no formal writing instruction and little usable guidance about what good writing looks like, writers may have to rely on what they did before, as Baker (2017) found with students continuing to use A-level criteria to guide writing. Schwartz et al. (2012) suggest that students require rich feedback in order to prompt noticing that a well-rehearsed routine is not appropriate for a new situation. They note that much feedback is of insufficient quality, however: feedback that is targeted at correctness may not illuminate to the learner that their familiar, comfortable approach isn't working out. Taking a self-regulated learning

perspective, monitoring learning to discover a well-worn approach isn't working is contingent on students' standards. Evaluative experiences such as peer assessment that enrich students' evaluative expertise and so standards may, then, help them to realise that they need to adapt their approach, and also provide possibilities for how they can enact this.

Focussing on the large-scale adaptations students make over time to their essays should be fruitful for examining writers' development. Much self-regulated learning research that is process-oriented rather than disposition-focussed examines adaptation within tasks. Tracing adaptations between tasks puts the focus of the study on the same level of granularity as most formative assessment studies, which look at the impact of interventions over time. The self-regulation dimension adds nuance. Formative assessment research tends to focus on judging learning by performances, for example curriculum-based assessments, or a writing test. Yet performances do not always reflect learning: an approach which leads to an immediate error-strewn performance may actually lead to better longer-term learning than an approach which yields a highly accurate immediate performance (Soderstrom and Bjork, 2015). In addition, performances such as an essay may not display the strategies that a writer has been trying out, which may not lead to improved performance immediately, anyway (Camp, 2012). Returning to the metacognitive definition of writing used in this study, a writer's words are at once both product and process, but a reader's insights into the process are limited (Hacker et al., 2009). Examining large-scale adaptations, with the writer's help, should give more nuanced insights into the strategic aspects of writing development, the need for which has been highlighted in the writing literatures (Alamargot and Fayol, 2009; Slomp, 2012).

3.10 Chapter conclusion

The theory of action brings together concepts from SRL, writing and formative assessment to model how evaluative experiences such as peer assessment might have an impact on self-regulated writing processes. By synthesising key theory and insights from these fields, the theory of action should enable a richer account of the impact of peer assessment on writing. The next chapter describes how the research design and methodology addresses the theory of action and research questions.

Chapter 4
Research Design and Methodology

4.1 Chapter introduction

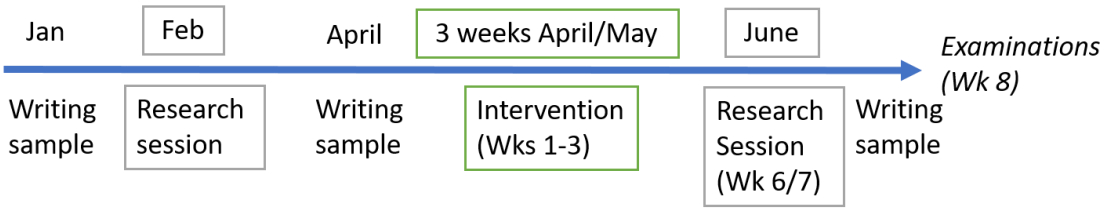
To address the research questions raised by the theory of action, the study used a design-based approach, and employed a range of methods. A key challenge in this study was to generate high quality research data while ensuring that participation was not overly burdensome to participants; above all, students were preparing for examinations. This chapter describes how the challenge was met. First, I will give an overview of the design and data collection, and outline how the chapter will progress.

As detailed in this chapter, the study drew on methods from the fields of writing processes and SRL to address the over-arching research question of whether and how students' self-regulation of writing develops in the context of peer assessment writing workshops. In the main study (2015), 21 students took part, 10 of whom serve as case study participants, and whose developing self-regulation of writing is examined in more depth. The study is small-scale in terms of participant numbers, but draws on rich and detailed material, with the main dataset totalling approximately 250,000 words. This includes participants' workshop data, and the pre- and post-workshops writing sessions with the case study writers. Figure 2 gives an overview and timeline for 2015 data collection. I met with case study participants in February for the pre-intervention research session. They wrote a timed examination essay, thinking aloud up until they wrote the first full sentence of the essay, and participated in a post-writing interview. The intervention workshops took place in April-May, over Weeks 1-3 of Trinity (Summer) term. Case study participants then met with me at their convenience for the post-intervention writing session in Week 6 or 7; the examinations were held in Week 8.

Participants also provided writing samples – from college collections exams, termly formative examinations set by tutors. Some students also submitted a practice essay, where written as part of their exam preparation.

The chapter opens by presenting the overall approach to the study: design-based research. I then give an overview of which methods were used to address each research question. Next, the setting for the research is introduced, and how decisions were made to suit the context of the study. I then describe sampling, followed by the design and procedure for the workshops, and how this was achieved. Following in-depth discussion of the methods, I outline the research ethics procedures, and discuss reflexivity. I then give a detailed account of the analytical strategy. The chapter concludes with a discussion of the limitations of the research design and methods.

Figure 2: Timeline and overview of data collection



Data from pre- and post-intervention research sessions:
- Think-aloud protocol – choosing question and planning (n = 10 in February; n = 5 in June)
- Timed essay
- Post-writing interview

Intervention data includes: annotations and responses to example essays, timed essay, peer feedback, written reflections, self-efficacy and background questionnaires

Writing samples: termly college exams (collections), and where written and provided, a pre-examination practice essay

4.2 Design-based research approach

The overall approach to the research is design research methodology, which aims to both build and test theory and contribute to the development of practice. It aims to take a middle ground between the generalisations of theory and the narrow specificity of findings limited to a particular context (Brown, 1992; Cobb et al., 2003). Key features of design-based research include: an interventionist approach situated in a 'real' setting; collaboration with practitioners; a focus on theory generation, testing and development but also implications for practice; multiple iterations and cycles, with design issues emerging and being refined at each stage; mixed methods evaluation intended to capture different aspects of the intervention's processes (Anderson and Shattuck, 2012; Cobb et al., 2003; McKenney and Reeves, 2013). Design research studies often start as small-scale (e.g. MacArthur and Lembo, 2009).

At present, the study consists of two initial design research iterations — 2015 and 2016 — with the possibility of scaling up and initiating another iteration after DPhil. The aim of this study is to find out whether the support for peer assessment for writing in the literature translates into empirical promise when used to address two practical problems. The first is limitations of feedback as the key way students learn about disciplinary writing, and the second the mis-alignment between tutorial and assessed genres of writing. The design of this study does not aim to show whether peer assessment 'works,' whether it provides a neat statistical solution to these problems (e.g. Slavin, 2004). Design research can test causal hypotheses, but such research questions only belong in the final stages of the cycle, once the issues that affect the way peer assessment unfolds have been clarified and once its promise has been demonstrated (Phillips, 2006). The early stages of the cycle are important, because if an intervention does not show promise in a small-scale implementation with researcher support, for example, in a case study, then it is highly unlikely to work on a larger scale (Brown, 1992; Graham, 2014). The

study uses a single-group design, appropriate for this early phase of design research (e.g. MacArthur and Philippakos, 2013). It would neither be practicable nor ethical to use a quasi-experimental design as comparison groups must offer worthwhile and credible alternatives (Graham and Harris, 2014). I am aware of the sorts of issues which may cloud single-group pre-post-test results (Marsden and Torgerson, 2012). The setting for the research – the first year of the Oxford English degree – is a critical case (Yin, 2014). If the peer assessment proposed is not practicable with high-achieving students who write frequently and are adapting to a new genre, then it is unlikely to work in other HE settings.

On the next page, Box 1 recaps the Research Questions, and Table 1 gives an overview of the methods used to address them. The methods generate data to examine both whole group and intra-individual development. On a group level, in 2015 and 2016, participants completed a pre- and post-intervention self-efficacy questionnaire; this was complemented by a writing processes questionnaire in 2016. A range of naturalistic data were collected: annotations and responses to example essays, feedback on peer essays, reflections written at the end of each workshop. In addition, I collected the timed essay written during the workshops, and participants were asked to provide other writing samples. To trace intra-individual development more closely, in 2015, ten case study participants met with me before and after the workshops for a think-aloud, timed writing and interview. As Valsiner (2000, p.107) points out, all data is socially constructed and 'data are signs that represent selected features of the phenomena' that is being studied. Thus, the methods were chosen to be consistent with the dynamic nature of self-regulated writing, as emphasised in the theory of action and research questions. It is acknowledged that the methods may have served to raise metacognitive awareness, which is a key aim of the intervention (e.g. Flower et al., 1990); and indeed some

Box 1: Research Questions

(How) does students' self-regulation of writing develop in the context of peer assessment writing workshops?

RQ1. During the workshops, which criteria do students use to evaluate written essay products?

RQ2. Which criteria do students use as standards to self-regulate their essay writing processes?

RQ3. (How) do students' task perceptions shift over the course of the study?

RQ4. (How) does peer feedback contribute to students' self-regulation of writing processes?

RQ5. What is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing?

Table 1: Overview of Research Questions and Methods

Key: W1 = Workshop 1, and so forth

	RQ1. Criteria	RQ2. Standards	RQ3. Task perceptions	RQ4. Feedback	RQ5. Self-efficacy
Pre-intervention					
Think-aloud protocol, timed essay, interview (case study participants only)		X	X		
Writing sample (college exams)			X		
During intervention					
Self-efficacy questionnaire (W1 & W4)				X	
Annotations and responses to example essays (W1, W2, W4)	X				X
Written reflections (W1, W2, W3, W4)	X	X	X		X
Timed essay			X		
Annotations and feedback on peer essays	X				X
Workshops evaluation questionnaire		X	X		X
Post-intervention					
Think-aloud protocol, timed essay, interview (case study participants only)		X	X		X
Writing sample (college exams)			X		

workshop activities were designed to do so. This is unavoidable. As Winne and Perry (2000) point out, every measurement of learning, whatever its form, is in some way an intervention. Even methods which may be assumed to have less of an impact, such as questionnaires, are interventions, and I reflect on the socially constructed nature of the data throughout this chapter.

As no single method in isolation can offer a rich enough view of the processes under study, methodological triangulation is used. Indeed, SRL and writing studies in this area are increasingly multi-method in design (e.g. Butler, 2011; Butler and Cartier, 2017; Myhill and Jones, 2007). The multiple data sources are used to build a richer, more illuminating picture, but without the assumption that this is unproblematic, or complete (Hammersley, 2008). Thus, the underpinning philosophical approach to the study is broadly pragmatic (Johnson and Onwuegbuzie, 2004; Robson, 2011). The data generated are predominantly qualitative not because I hold a particular commitment to qualitative research, but because the methods are best placed to answer the research questions. It is hoped that the methods and findings used will form the basis for measures used in larger-scale, future iterations of the study. As discussed earlier, a key aim of the 2016 iteration was to pilot an adaptation of the Regulation of Learning Questionnaire, examining its potential to generate rich data about self-regulated writing on a larger scale.

As part of the design-based approach, the study and intervention was designed to serve learning in the particular setting of the Oxford English degree. The workshops used authentic examples of writing, and interviews with tutors were very important to the selection of examples, as well as the design and planning of the workshops. The design was also responsive to issues that became important during data collection. For instance, following the pre-intervention interviews and tutor interviews, it was clear that confidence, or self-efficacy was

perceived to be particularly important to writing, and so self-efficacy was added as a research question, as discussed in Section 4.7.1.

4.3 Setting for the research

The setting for the research is the first year of the English Language and Literature course at Oxford. This case is a critical case as the tutorial system offers affordances for writing not present in other environments, and students have a clear transition from one genre to another (Yin, 2014).

Selecting such an ideal case is helpful in the initial stages of design-based research. If such a challenging form of peer assessment does not 'work' in favourable conditions then it is unlikely to in other contexts. The degree is divided into 'papers' or courses, and colleges retain their historic responsibility for teaching, offering tutorials and classes. The English Faculty sets the examinations and offers lectures and some special options teaching. Notably, there is no Faculty syllabus - college tutors develop their own courses and set reading lists which match their own expertise and interests (Horn, 2013). As a result, there is considerable variation between colleges in teaching and learning.

As is the norm at UK universities, writing is not explicitly taught as part of the English course. There is no writing centre, although eligible students can receive funded study skills support through the Disability Advisory Service. Students produce a significant amount of writing, typically up to 12 essays and commentaries over each 8 week term. Essays are the focus of discussion in tutorials between 1-3 students and a college tutor. Practices vary: traditionally, essays are read out, or they may be exchanged in advance. They may also be completed after the tutorial. Students often receive both verbal and written comments. The extent to which

such feedback focuses on writing will obviously vary between tutors and between colleges, depending on emphasis on writing, and interest and expertise. The tutorial essay genre is not, however, the main examined genre. While portfolio assessment of essays of a similar length has been introduced for one first year paper and one finals paper, most examinations consist of three essays written in three hours in response to a set of unseen quotations (3 of 4 papers for Prelims and 4 of 7 papers for Finals).

The intervention provided writing workshops for first year students preparing for Prelims Papers 3 and 4, traditional period papers spanning Literature in English 1830–1910 and 1910 – present day, respectively. These exam essays are part of the genre family of 'essay' (Gardner and Nesi, 2012) but differ both in the conditions of production and in the task demands to tutorial essays. Thus, the form of assessment is not aligned with the formative writing produced throughout the year (Biggs, 1996). Horn (2013) argues that such a 'constructive misalignment' affords tutors, and to a lesser extent, students to personalise papers and follow academic interests. However, as she points out, focus group research conducted by Oxford University Student Union (2010) found that students find this misalignment troubling, and this has been highlighted in other reports (e.g. OUSU, 2015). Although marking conventions and examiners' reports are available, in the National Student Survey 2016, only 55% of Oxford students and 63% of English students specifically agreed that 'the criteria used in marking have been clear in advance' (Oxford University, n.d.; Unistats, n.d.). Both percentages are much lower than for other items related to assessment and teaching, and also compares poorly to other universities. Students typically sit collections at the beginning of each term - college-set formative examinations on the previous term's work - but may find it hard to connect collections to examinations, as raised by the Student Union (OUSU, 2015). Collections are used to monitor progress, and students can be asked to sit 'crucial' collections if their progress is

not considered satisfactory by the college, or to re-enter the course after a period of absence (OUSU, 2015).

A notable issue which must be raised is that in English and some other subjects, a gender gap in achievement persists, with a greater proportion of male candidates receiving firsts than females. In 2016, of a cohort of 231 students, 72 candidates achieved Firsts in English Finals: 38 or 24.1% of female candidates and 34 or 46.6% of male candidates. For Preliminary or first-year examinations 57 of the 224 candidates achieved a Distinction. Again, a higher proportion of male candidates achieved the top grade: 42 or 33% of male candidates as compared to 58 or 22% of female candidates (ELL, 2016). The gender gap is often reported in the national (e.g. Eastham, 2014) and student media (e.g. Benhamou, 2010). The gender gap persists despite comparable verbal intelligence scores and school examination performance, and emerges in first year exams (Mellanby, Zimdars and Cortina-Borja, 2013).

While I have long been aware of the gender gap, the intervention was not designed to or framed as addressing this inequality. However, the majority of participants were female. In 2015, 17 of the 21 participants and in 2016 13 of the 14 participants were female. As these figures reflect the make-up of college intakes, little can be inferred. Indeed, in 2015, two male participants were the only representatives from their colleges. During the workshops, students were given a range of examples to appraise, rather than one ideal model which could be perceived as 'masculine'. I did not tell students the gender of the writer of the example essays, and we used 'he' and 'she' interchangeably. In the study, gender is seen as just one of the individual differences between students, rather than as binary variable to make generalisations. Individuality, diversity and the interaction with other differences are far more interesting in a study of this scale (Cameron, 2005; Jones, 2011).

4.4 Sampling and participants

In 2015, 21 students took part in the workshops; in the 2016 iteration, 14 participated. Most participants were aged 19. To take part in the study, students had to be first year undergraduates in English or joint honours at the University of Oxford. The workshops were offered to participants through college tutors. In January 2015, I was introduced via email to a sample of tutors by Kathleen Quinlan, Course Leader of the Postgraduate Diploma in Learning and Teaching in Higher Education. The rationale for purposively sampling diploma alumni was that they may be more interested in teaching and learning and more open to students' participation in the research. After the introduction, I emailed tutors with details of the project, and those who were interested and considered participation to be manageable for students' workload then offered the workshops directly to their students. Over the two years of the study, tutors from eight colleges invited students to participate in the workshops.

In 2015, students from seven colleges took part. The response rate varied from almost whole cohorts to just one student in the year group. In 2016, four colleges were represented. The response rate was lower in 2016 for several possible reasons. Firstly, I contacted college tutors much later than in 2015 and students were invited over the vacation rather than during term. At College D, one student was interested in taking part but was unable to attend the workshops due to illness. College E students were not invited due to a change in staff. In 2016, I met the College H tutor through a personal connection shortly before term started, and the students only had several days' notice.

At College A, the whole cohort took part in 2016. This was perhaps because the 2015 cohort did particularly well in Prelims, and while I was pleased for the students but cautious in

communication about the possible impact of the workshops, the tutors invited students particularly enthusiastically.

Table 2 gives the number of participants per college in 2015 and 2016. The total intake by college varies and is not given as this could risk identifying particular colleges.

Table 2: Participants by college		
	2015	2016
College A	7	9
College B	5	2
College C	1	1
College D	1	0
College E	1	-
College F	1	2
College G	5	0
College H	-	0
Total	**21**	**14**

The majority of participants in the workshops were women, as outlined in Table 3 below. In the three year period 2013-2015, 463 women and 235 men were admitted to study English. Women represented 66.3% of the cohort and men 33.6% (OU, 2015). Thus, in the study women are over-represented and men under-represented. This may reflect the factors in the institutional environment that leads to the gender gap in achievement in English, discussed elsewhere in this chapter. However, the gender imbalance is to an extent a result of college cohort imbalances. In both years, the College A intake was predominantly women. In 2015, 90% of the participants self-identified as White or British; in 2016, this figure was 64%. I have not gone into further detail here as it could risk identifying students from underrepresented groups. All identified English as their L1, although some also spoke other languages.

Table 3: Participants by gender		
Female	16	13
Male	5	1
Total	21	14

As shown in Table 4 below, most participants went to school in the UK. As detailed in Table 5, the majority attended state comprehensive schools. In the three year period 2013-2015, 366 students from state schools, 273 from private schools and 59 from 'other' schools were admitted to read English (OU, 2015). Thus students from state schools represented 52.4% of acceptances, and state school students are overrepresented in the study. Again, this could be partly explained by college intakes. In a further post-workshops questionnaire, all participants reported some experience with formative assessment activities, including peer assessment, prior to coming to Oxford.

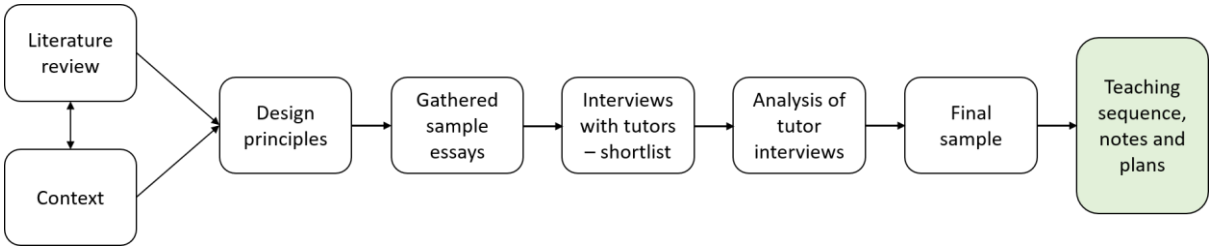
Table 4: Breakdown of pre-university schooling – countries			
	2015	2016	Total
UK	19	12	31
with some international, but UK for sixth form	1	2	4
International	2	2	4

Table 5: Breakdown of pre-university institution types			
	2015	2016	Total
UK state	14	9	23
- of which selective	(3)	2	5
UK independent	5	3	8
- of which selective	(5)	2	7
International	2	2	4

4.5 Designing the workshops

The process of designing the workshops is summarised in Figure 3 below. First, I developed a set of guiding principles from the research literatures that were appropriate to the particular disciplinary and institutional setting. These were: to review a range of authentic examples of writing; to conduct holistic assessment, with no pre-determined criteria; anonymity of peer assessment; comment-only assessment; opportunities for reflection to connect learning to own writing practices.

Figure 3: Designing the workshops



It is important that students evaluate a diverse range of writing, and that these pieces are authentic, that is not fabricated for teaching purposes (Sadler, 2010). To meet these principles, I assembled a small corpus of examination essays on a range of questions. The ideal source would have been Prelims examination scripts, but it was not possible to obtain these for various reasons. Instead, I gathered essays written for college collections (termly exams) and examination practice from second and third year students. These writers responded to a call for essays sent out by college tutors. The call emphasised that a variety of quality of work was sought. Students were asked to give informed written consent, and were informed that the corpus will be shared eventually with the Faculty for teaching purposes. Students were paid a small fee of £5 for their submission, as per studies such as the British Academic Written English Corpus (Nesi et al., 2008).

Seven students submitted a total of sixteen essays to the study. The essays were typed up and all identifying information removed. A two-step process was used to select essays for use in the workshops. Firstly, essays which were unlikely to have been handwritten in timed conditions were excluded. Where students submitted multiple essays, only one was included. The shortlist of essays was then discussed in interviews with four English tutors from participating colleges to develop teaching notes, and for two triangulation purposes: convergence and completeness (Gielen et al., 2011; Miller, 2003). These interviews allowed me to check that my impressions of quality and salient criteria were congruent with those held by tutors. They also allowed for completeness of criteria – each text was discussed with two tutors. These hour-long interviews mirrored the procedure used with students in the intervention, with tutors first asked for their overall impression of quality, before discussion turning to the key features which particularly contribute or detract from the quality of writing. The essays were also a starting point for discussion of common issues that undergraduate

writers face. As with the pre-intervention writing sessions, confidence, or self-efficacy arose as a key issue, with tutors perceiving that it had a considerable impact on writing and performance.

Due to time limitations, the tutor interviews were not fully transcribed; rather I pasted each essay onto a large piece of paper, and annotated the essay with transcribed tutor comments, using a different colour for each tutor. This enabled me to see convergence and divergence of judgement on each essay. Five essays were then selected to cover a range of quality, criteria, length and topic. The annotated essays were used to develop teaching notes on each essay, which I referred to during the group discussion to ensure coverage of all salient criteria. Appendix A shows this process for one essay and provides the resultant teaching notes. Next, I sequenced the essays to ensure that students encountered a range of quality from the first workshop. I then developed lesson plans, as described in the next section.

The five essays selected were intended to show a range of quality. I initially planned to have a pre- and post- test essay to capture shifts in students' comments over the workshops, and selected two similar essays for this reason. However, designing a reliable yet meaningful pre-post-test is challenging, and I reflected during the workshops that including a wider variety would be more beneficial to the students' learning. As a result, I selected as the final example essay a slightly more difficult and unusual essay which raised other issues, such as interpreting the question/quotation and dealing with multiple texts.

4.6 Procedure

As outlined in Table 6 below, the intervention was split over four sessions of up to two hours each, and took place in weeks 1-3 of Trinity term. Participants were split into three teaching groups in 2015 and two in 2016. Students were allocated to group according to their schedule, so college cohorts were taught together. The workshops took place at the Department of Education. The teaching rooms were set up with desks in a U-shape for ease of discussion. Water and biscuits were provided throughout. Handouts were designed so the data could be easily immediately anonymised and filed away securely. All handouts were printed on carbonless paper, instantly generating a research copy and student copy. Materials can be found in Appendix A.

In Workshop 1, informed consent was obtained and any questions about the workshops answered. After the participants had completed questionnaires, as described in the next sections, the workshops were introduced. Students introduced themselves to each other, and made name signs. The aim of the workshops was introduced as to develop ideas about quality in exam writing through holistic peer assessment, discussion and reflection on participants' own writing. The scope of the workshops was defined as involving comment-only assessment, so students were not asked to give grades or marks. It was explained that assigning a mark or classification is a separate skill that develops after developing ideas of quality, and is particularly challenging in a subject such as English. Participants were introduced to holistic assessment by contrasting it to what I termed 'criteria-first' assessment, as in a checklist or rubric. This 'top-down' approach was contrasted to a holistic, 'bottom-up' approach. Holistic assessment was characterised as considering which aspects of an essay stand out and contribute to the quality of the essay as a whole, either because they contribute to quality, or hold the essay back.

Table 6: Workshops procedure	
Workshop 1	Consent and discuss nature of study. Complete self-efficacy questionnaire (and RLQ in 2016) (10 mins). Overview of workshops and principles of peer assessment (10 mins). Individually, students read, annotate and write a response to Example Essay 1 (15 mins). Discuss as a group (15 mins). Repeat procedure for Example Essays 2 and 3. Discuss possible criteria, and note up on board (10 mins). Complete reflections (10 mins).
Workshop 2	Individually, students read, annotate and write a response to Example Essay 4 (15 mins). Discuss as a group (15 mins). Review criteria (10 mins). In timed conditions, write one essay from a past paper, period of choice (60 mins). Complete reflections on writing process (and RLQ in 2016; 15 mins).
Workshop 3	Discuss principles of feedback, and ground rules (10 mins). Give written feedback on three typed up and anonymised essays (up to 90 mins, at students' own pace). Complete reflections on feedback; re-read own essay and reflect (15 mins).
Workshop 4	Discuss receiving feedback (10 mins). Receive written feedback, plus typed up version of own essay. Reflect on feedback (15 mins). Individually, students read, annotate and write a response to Example Essay 5 (15 mins). Discuss as a group (15 mins). Discuss potential strategies for developing writing, and other groups' criteria. Share and discuss strategies handout (20-25 mins). Complete questionnaires (self-efficacy, RLQ in 2016, general reflections on workshop, previous experiences with formative assessment, demographics; 20-25 mins).
Note: the discussion of Example Essay 3 took less than the allocated time as it was much shorter and the weakest example.	

Students were given 15 minutes to read the essay, and write a response, annotating as they wished. Participants were asked to focus on *how* the student wrote about literature rather than *what* they wrote. A list of possible questions students could think about when writing a response was shown. These included: what stands out as being particularly successful about this essay? What, if anything, do you find confusing? We also discussed what constituted a

response. Students were asked to write about what they noticed, and to try to write in full sentences and explain what they meant. A response could be personal – they could use 'I' and write about what they liked and what they found confusing or hard to follow. Students were free to annotate the essays as they wished. It was emphasised that the response didn't have to be polished, and was not feedback to the original writer.

During the group discussion, students were asked to switch to a different coloured pen for any notes and annotations, referred to as 'additions' henceforth. I started the group discussion with an open question: 'what did you think about this essay?' It was important that students developed their own ideas about quality, so in my teacher role, I tried to make sure that the talk was based around student judgement, and student-to-student interaction. I did not want to be seen as an expert who had 'correct' answers, and with whom they had to agree; I wanted discussion to involve divergence as well as convergence (Pryor and Crossouard, 2008). In college cohort groups, the discussion was very fluid, with lots of student-to-student interaction. Where discussion faltered or interactions were directed back to me, I asked students what they thought of the previous judgement, or whether they agreed. I also nominated students who hadn't yet participated to share their views. However, as teacher, I did not take a laissez-faire or observer approach. Where discussion veered into subject content – for example, suggesting that the writer should have used a particular text – I guided it back to *how* the content was communicated. This was particularly the case in college groups. I also ensured that each group covered the key teaching points by ticking them off in my notes. Where points didn't arise in the discussion, I raised these first through questions. To prevent these from being too leading, I pointed to parts of the essay rather naming aspects – for example 'what did you think of Paragraph 2?' – rather than 'what do you think about the way the writer uses critics?' There were a few occurrences, with referencing and dates, where

issues didn't arise through this strategy, so I relayed to the students what tutors had said in interviews.

The procedure was repeated for Example Essays 2 and 3. At the end of each discussion, once students were working individually on example essays, I took brief fieldnotes on the nature and fluidity of the discussion. At the end of the Example Essay 3 discussion, I asked the group which aspects of writing had been important to the discussion, and if these were to be treated as criteria, what we would call them? Each group discussed the criteria and they were written up on the whiteboard. Finally, participants completed a reflections sheet.

Workshop 2 opened with Example Essay 4, and the same procedure was repeated. After the group discussion, we reviewed the list of criteria, and added or edited any as necessary. The participants then wrote an essay in timed conditions – 1 hour – on a paper of their choice. Participants could move to another spot in the room if they needed more space. They then completed a reflections exercise, as described in Section 4.7.

In Workshop 3, we began with a discussion on the differences between a response and feedback, which opened with an example of a comment. We discussed the differences, and how to channel a critical response into helpful feedback for the writer. I highlighted the difference between a deficit and a developmental perspective on writing, using a glass half-empty/full analogy. We set ground rules for peer feedback, such as using a constructive and helpful tone, and focussing on the aspects of writing which could really lift and improve the essay. The criteria developed by the group were displayed, but participants did not have to refer to them. Each participant received a pack with three typed-up, anonymised essays, and a feedback sheet. Participants read the essays and wrote feedback individually. This was at students' own pace. For most, assessing took around 45 minutes; one participant took 90

minutes. In 2015, as one participant was ill and was unsure whether he could attend the next session, I asked three students who finished early to give feedback on an extra essay. This student was eventually able to meet with me, so those three writers received four reviews. After writing feedback, students completed reflections, collecting a copy of their own essay part-way through to re-read their own writing. Participants could not keep the peer essays, but they could request a photocopy of their own comments to the writer.

Between Workshops 3 and 4, I typed up the feedback, including annotations where the reviewer indicated these were to be shared. In Workshop 4, we began with a discussion on receiving feedback. Following the issues discussed in Section 4.10, I decided that it would be helpful to discuss the importance of receiving feedback critically, and putting comments into context. Once they received their feedback pack – a copy of their typed up essay, and the comments – students read and annotated as they wished, and completed a reflections sheet. We then reviewed Example Essay 5, using the procedure as before. To round off the workshops, we discussed strategies for developing essay writing. After students had shared suggestions, I distributed a handout containing a synthesis of the criteria across the workshop groups, points from tutors which weren't often raised by students, and ideas for developing writing. Finally, participants completed questionnaires: on their previous experiences with formative assessment, to evaluate the workshops, and if they hadn't done so before, demographics.

4.7 Research methods: during the intervention

The workshops activities were designed so that participation would not be overly burdensome, and to generate a range of data, naturalistic and research-oriented. All workshop materials

were provided on carbonless paper, so students could take away an instant copy. I provided photocopies where the duplicate was unclear. The naturalistic data consisted of the annotations and responses students made to the example essays, along with the additions they made during discussion. I also collected the timed essay, the written feedback and annotations on peer essays, as well as any written comments and annotations made on the feedback students received. These data provide traces of students' developing evaluative expertise and which criteria were to become standards.

For both pedagogical and research purposes, participants completed 'reflections' at the end of each workshop – a set of 5 or 6 self-report questions intended to review learning and prompt metacognition about writing. These are included in Appendix A. In Workshop 2, reflections were written immediately after the timed essay, and were more research-oriented, with a total of 16 items about the writing process, to mirror the pre- and post-workshop interviews. Two participants commented that they found the reflections long and tiring, especially after writing a timed essay, and in 2016, participants responded to the modified RLQ instead, described below. This was intended to be less burdensome, with fewer qualitative items.

Participants also completed research-oriented questionnaires. The self-efficacy instrument and RLQ are described in the next section. In Workshop 4, participants completed a short questionnaire about their prior experience with formative assessment activities, both at Oxford and in their prior education. They also completed a short evaluation of the workshops, gathering their perceptions on workshop learning and the usefulness of the activities. Non-case study participants also completed a short questionnaire to provide background demographic information. These questionnaires are included in Appendix B.

4.7.1 Self-efficacy questionnaire

As confidence for exam writing was a recurrent theme in tutor interviews and the pre-intervention interviews, I decided to add a research question to address this. A short questionnaire administered at the beginning of Workshop 1 and then the end of Workshop 4 was considered to be the most time-efficient method, and an amendment to CUREC was obtained.

Existing questionnaires were either designed for school writing (e.g. Bruning et al., 2013) or were very general, also covering reading (Prat-Sala and Redford 2010, 2012). Thus, I designed a short genre-specific questionnaire to address the demands of the exam writing task. Items were written both to reflect aspects of writing described in the pre-intervention interviews, and to draw on Bruning et al.'s (2013) model of self-efficacy for writing, which proposes three dimensions: ideation, conventions and self-regulation. Following the principles from Bandura (2006), the survey was designed with 'I can…' statements and a 100 point scale. Due to time constraints, it was not possible to pilot the questionnaire, but I discussed the items with my supervisors and others. The final instrument consisted of 18 items.

In both iterations, all students completed the self-efficacy questionnaire at the beginning of Workshop 1 and end of Workshop 4. One student was excluded from the 2015 analysis due to illness. Statistical analysis of the 2015 results is described with the findings in Section 5.9. The instrument is included in Appendix C.

4.7.2 Regulation of Learning Survey (RLQ)

In 2016, participants completed a modified version Regulation of Learning (RLQ) survey, an instrument developed by Dr Allyson Hadwin and her lab at the University of Victoria (e.g.

McCardle and Hadwin, 2015). The original instrument asks respondents to reflect on and rate a particular study episode, so I adapted the items to make them appropriate for writing and the setting. The primary purpose of using the RLQ was to investigate the possibility of using the instrument in future research, and so the results are not reported in the thesis. The instrument is included in Appendix D.

In 2016, the RLQ was given on two occasions. Time 1 was the beginning of Workshop 1, prior to the self-efficacy questionnaire. All participants answered about their college collections (midterms) essays in Week 0, so around a week before Workshop 1 (Week 1). Time 2 was the end of Workshop 2 (Week 2 of term), immediately after finishing the timed essay. The survey was split into three sections, asking participants to reflect on their intentions for the writing, what they did before they started writing, and what they did during or shortly after writing. Students rated their intentions or actions on a 4 points scale. 40 items were used at Time 1, together with 6 open-ended items; at Time 2, 38 items were used as two were covered in the 10 open response items, which mirrored the Reflections used in 2015.

4.8 Research methods: pre-and post-intervention

In order to gather in-depth data on writing processes, twelve students took part in a pre-intervention writing session, and ten of these participants also completed the post-intervention session. These ten participants are the case study writers in this study. Attrition was due to illness. At the beginning of the sessions, we discussed the study and consent, and participants gave written informed consent. Immediately after each session and workshop, the data were anonymised and stored in a locked filing cabinet. Participants could choose their pseudonym if they wished.

The sessions were based around writing a practice essay and took approximately two hours. In order to capture task perceptions and planning, participants thought aloud up until the point where they started writing the first sentence of their essay. A post-writing interview focussed on writing processes, and participants' beliefs about quality and academic writing more broadly. The protocol for the writing session is included in Appendix E.

The students who submitted essays for the workshops were invited to participate in a pilot, but none were available. As a result, the first three cases were used as a pilot. As detailed in this section, interview items were added after the first session, with Nicole. After the second and third sessions, with Phoenix and Sean, the practice for the TAP was changed. To inform these decisions, I fully transcribed the TAPs and made notes and part-transcriptions of the interview recordings and compared them to my research questions and the theory of action. These were more satisfactory and no further changes. The cases of Nicole and Phoenix are included in the analysis as their data is largely consistent with the other participants; Sean did not attend the post-intervention session, and so is not a case study participant.

4.8.1 Think-aloud protocol and timed essay

As task perception is a particularly important process to writing, before participants started to write, they were asked to think-aloud. Think-aloud protocols (TAPs) give insights into (meta)cognitive processes which can only be inferred from behavioural measures such as keystroke logging, and have been a key writing research method since Hayes and Flower's seminal (1980) study and for studies on task perceptions specifically (Flower et al., 1990; Van den Bergh and Rijlaarsdam, 2001; Plakans, 2010). In another scenario, pairing think-aloud with keystroke logging would be useful for further methodological triangulation of processes

(Leijten and Van Waes, 2013), but the students in this study handwrote the essays, as they would in an exam. A key advantage of TAPs is that they do not rely on memory (Greene, Robertson, and Costa, 2011). The method has also become more popular in SRL research in recent years with the shift in conceptualisation of SRL from static aptitude to dynamic process (Greene et al., 2011). TAPs do not interfere with cognition so long as participants are asked simply to verbalise rather than to explain (Bannert and Mengelkamp, 2008; Ericsson and Simon, 1993; Greene et al., 2011). However, it cannot be assumed that TAPs directly convey thought – as with all data, they are socially constructed and participants will inevitably select which thoughts to verbalise.

Participants were asked to think aloud from opening the examination paper until starting the first sentence of the essay. This typically involved reading all or part of the questions, making a choice and planning. Participants were asked to stop verbalising at the first full sentence of the essay as thinking aloud while writing can be difficult, and recent writing studies have typically asked writers to verbalise only for short writing tasks (Beck, Llosa, and Fredrick, 2013; de Milliano et al., 2012). The mean duration for the case study participants' pre-intervention TAPs was 13 mins 46 seconds (range: 7:40 - 25:51). At post-workshops, the TAP was optional (see Section 4.10 for further discussion). Five participants opted to think aloud, and the mean duration was 9 mins 36 seconds (range: 6:26 - 15:09). The mean word count of the transcriptions was 1219 words at pre-intervention and 1036 words at post-intervention (excluding the practice).

Before participants were given the examination paper, they were asked to complete a practice task, for at least three minutes or until they seemed comfortable with the task. The mean duration of the practice was 5 mins 5 seconds (range: 3:05 - 7:00). Following Ericsson and Simon (1993), the pilot included word conundrums in the style of the TV show Countdown, in

which contestants try to make words out a random selection of letters. However, it emerged in piloting that participants were not used to such conundrums, and found the task difficult. As a result, the remaining participants were asked to annotate a poem, as this is a common literary practice. The poem, Roger McGough's 'Catching up on Sleep' (McGough, 2006), was chosen as it is contemporary, straightforward in lexis but had plenty of interesting features on which to comment. One participant said she found annotating and unpacking the poem reminded her of her admissions interview. After practising, I checked understanding and answered any questions, for example about what should be verbalised, and the difference between explaining and verbalising.

For the post-intervention think-aloud, participants were given the option of practising, and chose to use another past paper rather than a poem. Only two of the five participants chose to practice, and the durations were 2:25 and 3:08. I had not previously used past papers as I was concerned participants might be thinking of the practice items when going through the 'real paper' but this seemed easier as a warm-up as the task had a clear purpose.

Two interview items were used to check the perceived impact of thinking aloud on participants' usual approach. Many participants found thinking aloud 'weird' and a bit difficult, but few reported that it hindered writing. One mentioned that thinking aloud forced her to read all of the questions on the paper, indicating that the method raised her awareness of what she 'should' do with the paper.

4.8.2 Post-writing interviews

The post-writing interviews were semi-structured, enabling cross-case analysis but also leaving scope for follow-up questions to clarify, capture idiosyncrasies, and remain open to unanticipated issues around writing (Miles, Huberman and Saldaña, 2013). To ensure consistency, I ticked off items as I asked them. The 21 interview items were written to address the research questions, and emerged from the theory of action. The interview protocol began with an open question, 'how did it go?' This tended to prompt self-evaluation, but also gave participants an opportunity to voice any burning issues. The interview then focussed on the process of writing the essay, using mostly open-ended questions such as 'Tell me about how you came to choose this question'. The discussion then transitioned to task perceptions, asking participants about the question they had answered. Next, quality, asking participants to first compare the essay they had just written with the collections piece they brought to the session. They were then asked about which strategies from A-level or their prior education they could apply to the present task. The interview closed by asking the participant about the potential impact of thinking aloud, and anything else they'd like to add. By basing the interview around students' immediate writing, it was expected that students would talk more precisely about processes, and be more likely to use their own terms. More general interviewing about writing can lead to confused meanings and certain questions framing writing in a particular discourse, as Mitchell et al. (2008) found in their study of argument.

The mean duration of Interview 1 was 31 minutes 7 seconds (range: 21:19 - 48:28), with a mean word count of 6,202 (range: 3,515 - 11,399). At post-intervention, I added a further 4 items, which asked participants to reflect on what they had learned in the workshops, and what they had put into practice. As a result of this, and perhaps because participants were more used to talking about their writing, the interviews were longer. The mean duration of

Interview 2 was 43 minutes 25 seconds (range: 31:09 - 52:02) and the mean word count was 8,465 words (range: 7,151 - 11,466).

4.9 Reflexivity

Reflexivity is important to the research process and to the quality of the resulting research. Throughout the study, I have tried to practice 'uncomfortable' reflexivity, to use Pillow's (2003) term. This is reflexivity not for reflexivity's sake, or reflexivity that offers reassurance to the researcher, but reflexivity that aims to be unsettling. Throughout data collection and analysis, I have kept reflective notebooks and tried to question my assumptions. I have tried particularly to be reflexive with respect to the range of roles and identities I hold in relation to the participants: researcher; teacher; a fellow, but more senior student; a former A-level teacher; a former Oxford undergraduate; someone both inside and outside the discipline of "English"; a writing 'expert'. These shape how the data has been constructed and analysed. The most prominent roles that I needed to be aware of and balance were researcher and teacher of the workshops. I needed to make sure the students got the most out of the learning in the least unduly stressful way, with generating good quality research data - quality in research is also an ethical issue. For the most part, there was no conflict between these roles. The workshop activities were designed to support students' reflection and self-regulated learning, and at the same generate data on these processes which could be useful to both myself and the students. There were some activities which could be considered 'research-first' - the questionnaires, and I tried to design these to be a short as possible. However, one item did spark debate - a 2016 participant asked whether we could discuss strategies for emotional regulation in the final workshop.

As discussed in the analytical strategy section, due to my disability, I worked on much of the coding with a practical support assistant, Nadri Boswell. In total, 100 hours of support were funded and predominantly used on NVivo work, during which I directed Nadri to use the mouse on my behalf. This was both extremely helpful and quite demanding, as I had to verbalise and explain the coding as well as make coding decisions. To mitigate against fatigue, I prepared by hand annotating data in advance. The data and coding sometimes prompted discussions, which were helpful for bringing to light and examining my assumptions. As Nadri became more familiar with the coding guides, her occasional questioning of decisions was helpful for explaining the difference between codes, and helped me reflect on whether the codes were necessary. After we had coded the data, I ran rounds of checks on my own to ensure that the coding was consistent, and my judgements hadn't shifted. Thus, the practical support also prompted reflexivity.

4.10 Ethics

This section is based around Guillemin and Gillam's (2004) helpful distinction between planned ethical procedures and ethically important moments that arise during research. The section first addresses procedural ethics.

4.10.1 Procedural ethics: CUREC application

Throughout the study, informed and written consent was obtained from all participants. An initial CUREC application was submitted and approved in early 2015. This is included in Appendix F. The application covered workshop participation, and pre- and post-intervention

writing sessions, the students providing example essays and the tutors participating in interviews. Amendments were submitted and approved for the self-efficacy survey, and for the 2016 workshops. In January 2017, the Departmental and Divisional Research Ethics Committees agreed that it was acceptable for the practical support assistant, Nadri Boswell, to work with me on the NVivo coding under the terms of the original consent. Nadri signed a confidentiality agreement and did not access the data without me. Throughout the study, I followed the BERA (2011) and ESRC (2012) codes on research ethics.

In all cases, participants were sent the full CUREC information and consent form by email in advance of the first research encounter and asked to read it and contact me with any questions. When we met, they were given paper copies of the information letter and consent form, and given the opportunity to re-read and ask questions before completing the consent form. Very few participants asked questions, and where questions were asked they tended to be around procedural aspects of the workshops, such as who would read the essays. Throughout the research, I was very aware of the need to seek ongoing consent from students. I reflect on moments where this issue surfaced in Appendix G.

4.10.2 Ethical issues

Two key ethical issues accounted for in design and implementation were workload and attendance. Ideally, in a non-research setting or in future cycles of this design-based research the workshops would be integrated into regular teaching. However, this was not feasible for the present study, so participants had make a substantial extra time commitment of 7.5 - 8 hours over 2-3 weeks to attend the workshops, with an additional commitment of around 4 hours in total for case study participants. I was very conscious of this potential burden on

participants. Students at Oxford typically have a very heavy workload, which the Student Union regularly raises as a significant concern for welfare in numerous reports (e.g. OUSU 2015, 2016). Teaching over the three eight week terms is organised by colleges, and varies in nature, amount and scheduling but typically includes tutorials and classes, with additional lectures at the Faculty and significant independent study and research. Workload can vary across colleges and the 2016 QAA review recommended that the University address this by setting course norms. Having sat my own Prelims in English and Modern Languages at St Peter's College in 2002, I empathise with students and took careful consideration of the impact on workload. As discussed earlier in this chapter, survey data indicates that students find criteria and expectations for examinations opaque. The workshops were designed to address this, and I hoped that the time commitment would be offset by potential benefits.

The workshops were timed to fit around students' busiest points in term. First-year students write two portfolio essays in Weeks 4 and 5 and sit their Prelims examinations in Week 8. The workshops were timetabled for Weeks 1-3 of term to finish before the portfolio questions were released and to ensure that students had time to develop their writing after the workshops and the portfolio.

In order to inform decisions about participation, the time commitment was made clear in emails sent to tutors and students. In 2015, one college tutor decided not to offer the workshops for workload reasons, and others stressed that participation would be optional, which of course I agreed with. In 2016, no responses were received from two colleges who invited students, and I speculate that this was due to workload, particularly as the emails were sent in the last weeks of vacation. For one college, students probably didn't feel participation was necessary because the tutor already offered writing support and embedded peer support into teaching.

In designing the workshops, learning was the priority. To make the most of the time, the majority of the data collection was through collecting copies of the work students completed - annotations and responses, practice essay, reflections, feedback. Data assumed to be little direct benefit to students were: the self-efficacy questionnaire in the first and final workshops, and some items on the post-writing reflections in Workshop 2 which captured details about the writing process. In total, research-first activities accounted for around 40 minutes of the 7.5 hours of workshop time. In 2016, this increased to 45 minutes with the introduction of the modified-RLQ in workshop 1. However, the burden in Workshop 2 was reduced, as two 2015 participants commented on being fatigued after writing. The post-writing items were reduced to ten, and students completed the modified RLQ, which was hopefully less burdensome.

It was made clear to students that they did not need to prepare for the workshops themselves. However, most reported some sort of revision for the practice essay. As most colleges give 'collections' (formative examinations) at the beginning of each term, many students reported going over their collections revision, as I assumed would be the case, rather than doing fresh extensive revision.

Overall, it seems that the time burden was not perceived to be too onerous by most participants. In the workshop evaluation questionnaire, only two students mentioned time. These were 2016 writers from colleges where participation was low. No comments were made by students from colleges where most or all students took part; perhaps students saw the workshops as part of ordinary teaching, and could see how they could continue to use the learning.

With regard to the time commitment for the pre- and post-workshops sessions with case study writers, each session was limited to around two hours. Students who took part in these

sessions could choose which paper to write on, and so used the post-workshop session as part of their examination preparation. In addition, students were offered a small fee for participation in activities which were not part of the intervention workshops in recognition of their time and contribution. All participants who were offered payment accepted to do so and completed a receipt form which was submitted to finance for reimbursement from my ESRC Research Training and Support Grant. Students who submitted an essay for the corpus were paid £5. Those who participated in the think-aloud, writing and interview, lasting approximately 2 hours, were offered £20. Aside from workload burden, payment was offered for two main reasons. Firstly, payment for research participation seems to be the norm in Oxford (e.g. Warikoo and Fuhr, 2014). Secondly, by offering a small payment, it was hoped that students who struggle with writing or who do not enjoy the process would also be motivated to take part in the study, thus leading to a more diverse sample. I gave students the payment at the beginning of the research encounter in order to make it clear that was for for participation and not for their answers (Head, 2009). I also emphasised that the payment did not affect their rights as research participants.

The second key ethical issue was attendance. The workshops were timetabled in consultation with college tutors to minimise the potential for clashes. All participants completed the workshops in some form, but where students had other important commitments or were ill, I took a flexible approach. In 2015, one two participants left Workshop 2 fifteen minutes early and completed the reflections work out of class. A further participant attended the first workshop but was ill thereafter. I arranged with him to catch up outside of sessions. In case he wasn't well enough to write feedback, I asked participant from another group to review one additional essay. In the end, he was able to meet me to catch up, and completed the reviews and final workshop on a one-to-one basis. I decided to share all four pieces of feedback with

the three recipients - the feedback had been written and could be useful. Apart from the feedback he wrote and received, this student's data has been excluded from the analysis in the study as he didn't receive the intervention in a group setting. For the data for other participants, I have noted where they were not in workshops and take this into account during analysis. In 2016, two participants left Workshop 1 an hour early, one participant left Workshop 2 an hour early, and one participant missed the final workshop altogether due to illness. In the case of Workshops 1 and 4, I met with the students individually to discuss the example essays they had missed. The participant who missed part of Workshop 2 wrote the timed essay later that day, typed it up, and later provided the handwritten copy.

Handling ethical issues, then, was a matter of both planning and dealing with issues as they arose during the workshops. Further ethically important moments – situations which were impossible to plan for or anticipate – are discussed in detail in Appendix G. Such moments arose around feedback, ongoing informed consent, and anonymity.

4.10.3 Sharing findings with participants

It is important to me that research participants benefit from participation in the research. In 2015, I shared think-aloud protocol transcripts with those participants who were interested. Im addition, I had hoped that in the 2016-2017 academic year I could share research findings and theory that could be helpful to 2015 participants' Finals prepation. However, due to the delays to my DPhil for health reasons, this was not possible. I emailed participants to explain this, and offered to share a report on the research once finished.

A key outcome of the research will be to provide a resource pack for tutors at Oxford, including the example essays. I produced a draft resources pack in April 2017 and shared it

with those interested, but it did not contain the example essays as they cannot be shared until the degree is complete. I plan to update this when possible, and also to run a seminar for tutors.

4.11 Analytical strategy

This section of the chapter details the analytical methods applied to the qualitative data to answer the research questions: analytical memos, both data-driven and theory-driven coding and visualisations. Statistical analyses of the self-efficacy questionnaire and the quantified data are reported with the respective findings. The section then summarises which approaches were used to address the individual research questions.

The main dataset, as shown in Table 7, totals approximately 250,000 words. NVivo 10 is the main tool used to manage the analytical process, with spreadsheets also employed to track progress and collate word counts. While many researchers champion the use of qualitative data analysis software, I maintain a slightly sceptical stance and do not see NVivo as sufficient for conducting analysis and making interpretations, particularly when interpreting data at the individual case level. This section aims to explain *how* coding is used in the thesis, and not just what coding is conducted. While reporting frequencies is helpful for mapping whole-group trends, I want to avoid a reductionist approach, and coding alone is not sufficient for mapping individuals' development.

Table 7: Overview of main qualitative dataset		
Time	**Data type**	**Word count**
Pre-	Think-aloud Protocol	12191
	Essay	9029
	Interview	62018
During	Annotations, responses, additions	20505
	Essay	19050
	Reflections	17033
	Feedback	10419
	Annotations on feedback received	1430
Post-	Think-aloud Protocol	5180
	Essay	10076
	Interview	84652
	Total	**251583**

Despite its limitations, using NVivo is necessary for such a large dataset – for data management, easy retrieval and searching for particular items, and for running queries. Having used NVivo in a previous project (Usher, 2013), I was aware that I might find myself feeling detached from the data when analysing on-screen. This issue is noted in the wider methods literature, e.g. Luker (2008). In the previous project, using NVivo early on in analysis led to a proliferation of codes, which then had to be refined. These codes seemed to be taking the place of annotations, as the related software function is unintuitive. In this project, before any data was imported into NVivo, at least three cases of its type had been hand annotated and memoed so that I had an idea of the core codes that might come up. For case study participants, analytical memos were started pre-NVivo. Throughout my use of NVivo, I saved a new version of the project every time changes were made to coding schemes or data management, and kept notes on this. Despite taking these precautions, the software was not entirely reliable and I had to make manual checks throughout the analytical process. Due to my elbow and wrist problems, it was also physically very uncomfortable to use. As it should be

evident throughout this section, NVivo use prompted regular reflection, particularly when I was faced with its flaws and limitations. In their systematic review – itself conducted in NVivo – Woods, Macklin and Lewis (2016) propose that NVivo can enhance researcher reflexivity. In my case, most NVivo-prompted reflexivity occurred when I bumped up against the software's flaws and limitations, and hand coding generated the more valuable reflexive insights about my own position or about the co-authoring of the data, to use Brinkmann and Kvale's (2015) term. Throughout the analytical process, reflective notebooks have been kept to track emerging themes and assumptions. Indeed, analysis began with fieldnotes kept during the data collection process.

I will now outline how the analysis has been conducted. On the next page, Box 2 gives a timeline summarising the analytical process.

4.11.1 Transcription and preparation

After data collection, the first step of the analysis was to transcribe the data (Miles et al., 2013). All handwritten and audio data were transcribed verbatim, with pauses, laughter and sighs included for the audio data. This was to capture some of the social interaction in which the data were co-authored, and also traces of affect (Brinkmann and Kvale, 2015). For the handwritten data, close attention was paid to paragraphing, indenting and personal punctuation styles – e.g. preference for single or double quotation marks. As I needed the pre-intervention interview transcriptions for workshop planning, these data were transcribed by a professional under a confidentiality agreement, and then checked by myself.

Box 2: Timeline for analysis

February – June 2015: Transcribe pre-intervention TAP and interview data. Start reflective notebooks. Fieldnotes during data collection and intervention.

Summer 2015 – Transcribe intervention and post-intervention data; start analytical memos. Run self-efficacy statistical analyses; write up results.

Autumn 2015 – Develop analytical memos. Present at AEA-Europe (Glasgow). Initial feedback coding – present at SRHE.

Winter 2015 – Code responses, annotations and additions for criteria.

Spring 2016 – Overseas Institutional Visit to Dr Allyson Hadwin's lab, University of Victoria, Canada. Work on task perceptions. Start case study visualisations, adding to analytical memos.

April – May 2016: Run second iteration of workshops. Take fieldnotes.

Summer 2016 – Continue with visualisations. Use these to present at EARLI Writing SIG. Initial RLQ analysis.

Autumn 2016 – Visualisations complete, attempt to code in NVivo for criteria. Use visualisation data to present at AEA-Europe (Limassol).

January – April 2017 – Code case study data for criteria and SRL with practical support assistant. Refine coding guides and ensure consistency.

May – June 2017 – Run statistical analysis on criteria and standards. Code workshop data for criteria with assistant. Use a mix of data to present at EATAW.

July – September 2017 – Code feedback data for criteria with assistant.

October – November 2017 – Use NVivo queries to update analytical memos to trace large-scale adaptation.

Templates were used for the transcriptions to ensure consistency for each type. All transcriptions were hand checked, with the typed draft compared to the audio file/handwritten data. Interview, think-aloud and workshop data were prepared for NVivo, using heading styles to facilitate autocoding (Bazeley and Jackson, 2013). Autocoding was applied so all responses to particular interview questions or essays could be easily retrieved.

As with all other steps of the process, notes were kept to capture any interesting interpretations or details that arose during transcription. This was especially important to

capture some nuances of the interaction, as arguably, the spoken word loses its 'presence' once transcribed (St. Pierre and Jackson, 2014: 716).

4.11.2 Analytical memos

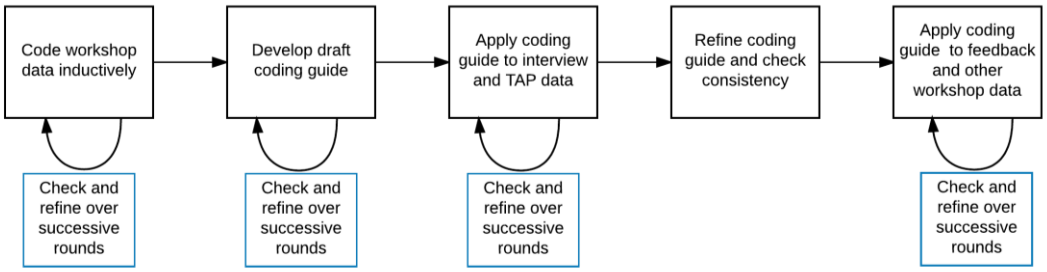
In the second step, I familiarised myself with the case study data, reading and annotating all data chronologically on a case-by-case basis. I then wrote analytical memos for each case study participant, describing shifts and changes and reflecting on key parts of the data, and in particular participants' evaluative talk about the own writing and what makes a good Prelims exam essay. The typical word count for the initial memos was around 5,000 words. These documents were living documents, updated throughout the research process, and now typically total 8,000 - 10,000 words. An extract from a memo on Megan's large-scale adaptation can be found in Appendix H.

4.11.3 Data-driven coding: criteria

The third step was to develop a coding guide to capture the criteria students use in their evaluation of essays. As detailed below, the results of this coding guide are used to inform the analysis for all research questions, and in particular RQ1. This coding commenced in Winter 2015. The approach to coding was inductive, or data-driven (Miles et al., 2013). The guiding question throughout the coding process was: '*What* is the participant commenting on in this response / annotation?' Put together, these comments reflect the underlying criteria students were drawing on as they commented on their own or others' writing. As the aim of the workshops was to enable participants to develop their own ideas about quality, it was not

appropriate to use existing sets of criteria such as the English Faculty criteria to develop the guide. Workshop groups developed their own lists of criteria, but as these varied between groups, it was more appropriate to develop a coding guide out of the themes identified from the data. Some codes were named in vivo, which as Reiff and Bawarshi (2011) found in their study of genre, was essential for capturing nuances in students' talk about writing. The coding guide is included in Appendix I. The process for developing this coding guide is summarised in Figure 4 below.

Figure 4: Development of criteria coding guide



The first type of data to be coded for criteria was the responses students wrote to evaluate example essays in the workshops. Initially, I tried to create a coding guide from the analytical memos and apply this to the responses to example essays. However, this approach was unsuitable as the memos are based on multiple sources of data and are richer and better explained than workshop data. Thus it was difficult to crystallise talk into codes that were succinct and not too idiosyncratic. As I was to start coding with the students' responses to the five example essays they evaluated in the workshops, I went back to this data exclusively. I read and annotated paper copies of the data, and then coded inductively in NVivo, including many in-vivo terms which recurred and indicated a possible shared discourse about writing. I began coding with three cases whom I knew less about, one from each workshop group. After coding these responses inductively, I reviewed the codes and wrote up an initial guide. I then

coded across the other non-case study participants, adding new codes only where there had been multiple instances across cases. The guide was then reviewed before I applied it to the case study participants' data, but I didn't remove any codes at this point.

After coding the case study participants' data, I reviewed the coding guide again. I exported the nodes into Excel and revisited in particular nodes which had fewer than five references, or three sources. I also considered whether some nodes could be subsumed into others. However, I didn't merge many nodes because at this stage I wanted the coding to remain open to further readings of the data. After rationalising the codes, I revisited the coding guide, and then carried out another round of coding for all the responses to ensure consistency. The coding guide was then applied to the annotations that participants made on example essays, and it was not necessary to further amend the scheme. At this point, there were 35 codes in the guide.

The minimum unit of analysis was one sentence clause. I coded chunks at multiple codes where necessary, rather than aiming for exclusivity. This was intended to allow exploration of relationships between criteria through later running matrix coding queries. It would also not be meaningful to separate out or judge which criterion was more dominant in one utterance. During this time, I began to develop a coding scheme to move beyond describing *what* participants were commenting on and identify *how* they were commenting. However, after coding a tranche of data, it was later decided that this wasn't necessary to answer my research questions.

Coding for criteria was put on hold in Spring 2016, when I visited the University of Victoria and started to focus on case study participants' self-regulation of writing, and developing visualisations. In autumn 2016, after completing the visualisations detailed below, I started to

apply the criteria coding guide to the interview and think-aloud data. However, I was unable to continue as NVivo use is extremely provocative to my condition as it requires heavy mouse use and does not offer integration with voice recognition software. Coding resumed in January 2017 with the help of the Practical Support Assistant, Nadri Boswell. We sat together, and I directed her to select and code text on my behalf. For much of the coding, the decision making was too complex to be done on the spot, so I hand coded before we met, scribbling codes on printed copies. Funding for the practical support was arranged by the university's Disability Advisory Service.

In early 2017, the interview and think-aloud data were coded according to the existing criteria guide, but this was treated as a work in progress, rather than a fixed artefact. Annotations were made in NVivo where data did not fit neatly, and definitions of codes were refined as necessary after the initial round of coding. The following new codes emerged during the first round of coding: terminology, finishing, range and scope, specific and selective, style – concise and precise. In the second round, the coding was checked, and new codes were applied consistently across the data set.

Next, the interview and think-aloud data were coded for SRL (see below). Once this was complete, the criteria coding guide was then refined to ensure that there were clear distinctions between codes, and that coding was consistent. For instance, the code *consistency and coherence* was added at this point to capture *focus* and to take some references coded under *signposting…* which were more to do with the reader's impression than how that impression was created. Codes such as *performance* and *time and process* were found not to be criteria as such, and so were moved to a separate coding guide – 'other', which covers references not to do with criteria or SRL. At the end of this process, the number of codes was reduced from 44 to 34.

The coding guide was then applied to the remaining workshop data and to the feedback data. Although I was open to new codes emerging, the coding guide did not evolve further. External quality checks such as interrater reliability were not relevant during the development of the coding guide, and have been beyond the scope of the project in the final months of analysis and writing. Nonetheless, the rigour of the process is assured through rounds of checks on the coding, and, later in the process, discussions with Nadri as she occasionally questioned decisions.

4.11.4 Visualisations and SRL coding

NVivo use breaks data down into codes and so fragments the data, making it hard to understand cases holistically and follow self-regulation over the study. During my visit to Dr Allyson Hadwin's lab at the University of Victoria, I developed a visualisation method to capture case study participants' self-regulation of writing over the study. The final version of this method, applied to all case study participants involves mapping interview transcripts, think-aloud protocols, workshop data onto phases and aspects of Winne and Hadwin (1998, 2008) and the theory of action: conditions, task perceptions, plans, enactment, monitoring, self-evaluation, adaptation. This start list of theoretical codes also reflects the interview questions, and the approach was overall primarily deductive (Miles et al., 2013). After the first case, a definition for each phase was written, combining the theory with the context. These are included in the SRL coding guide, described below. The visualisation process was carried out by hand, with printouts cut up and arranged onto a large sheet of paper under the different phases and aspects of learning. The data were then re-read in phase order, and

connections between phases drawn on the paper. An example of this process is given in Figure 5 for Chloe's self-regulation of writing.

Once this was completed for all three episodes of writing, a condensed pencil sketch tracing key aspects of self-regulation over the study was produced. Finally, this was digitised, currently with MS PowerPoint. An example, again for Chloe, is given in Figure 6, on the next page. Throughout this process, I added to my reflective notebooks. At the end of each visualisation process, I wrote up an overview of each writer's self-regulation over the study and appended this to the analytical memo for each case. As is challenging to make the visualisations reader-friendly without losing detail, I have not included them as illustrations in the thesis, but they have been helpful analytical tools.

While conducting this analysis, the importance of standards became apparent. Traces of standards were annotated on the visualisations. A second round of deductive coding on the large pieces of paper was later conducted to systematically identify the standards participants used to monitor and control their writing as a precursor to the SRL coding conducted in NVivo, discussed below. The visualisations and NVivo coding were cross-referenced in order to make the final selection of cases for discussion in the thesis, and to check the final visualisations.

The coding guide is included in Appendix J and the process for developing the guide is summarised in Figure 7 below. The coding guide is primarily theory-driven and was developed from the hand coding conducted for the case visualisations. I examined the suitability of published coding guides which had used Winne and Hadwin (1998), notably Greene et al. (2012). However, as this is the first study to apply the theory to writing, and also doesn't use an online platform, it was more suitable to develop my own guide from the interaction between the data and the theory.

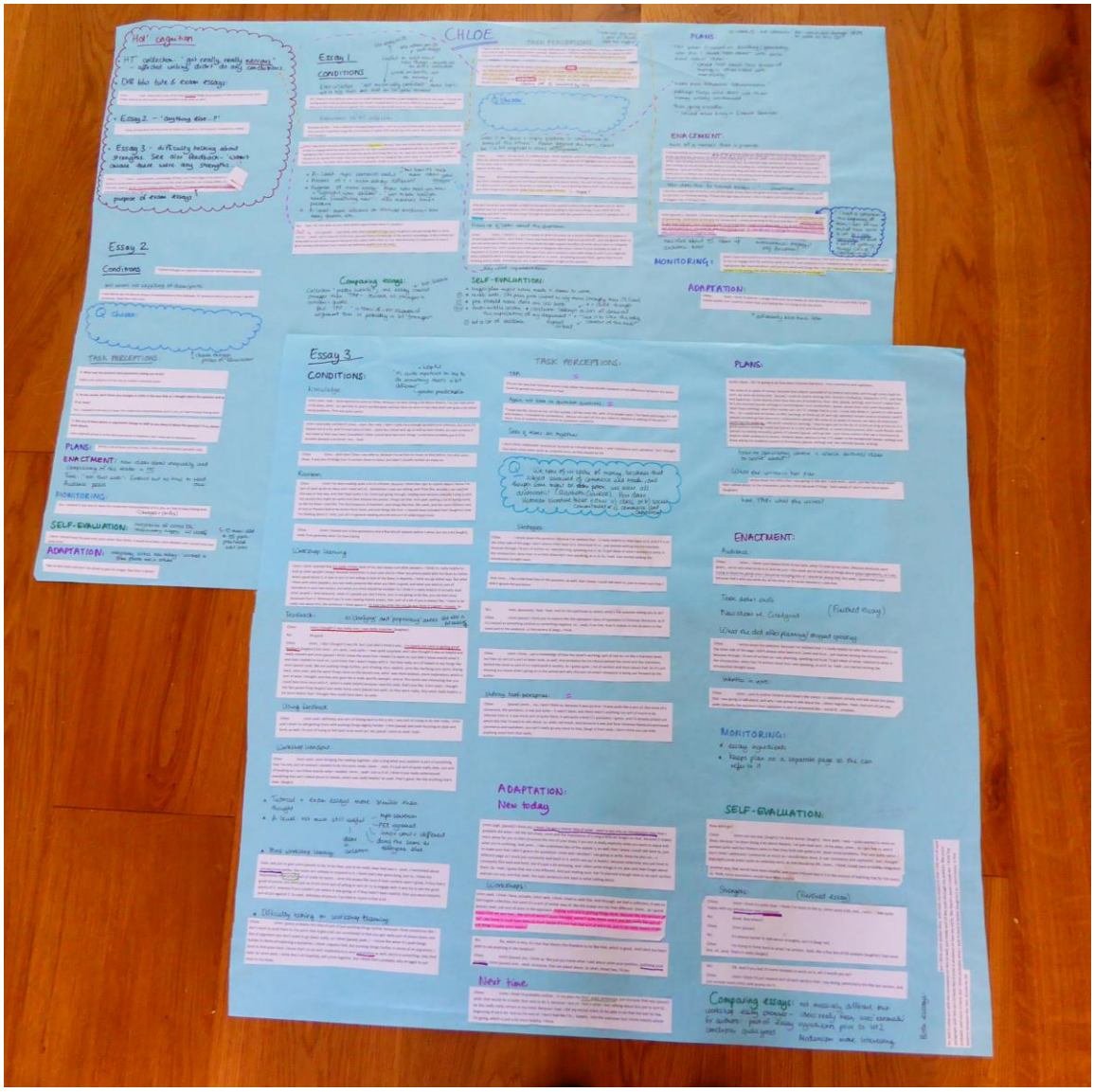


Figure 5: Example of visualisation process (Chloe)

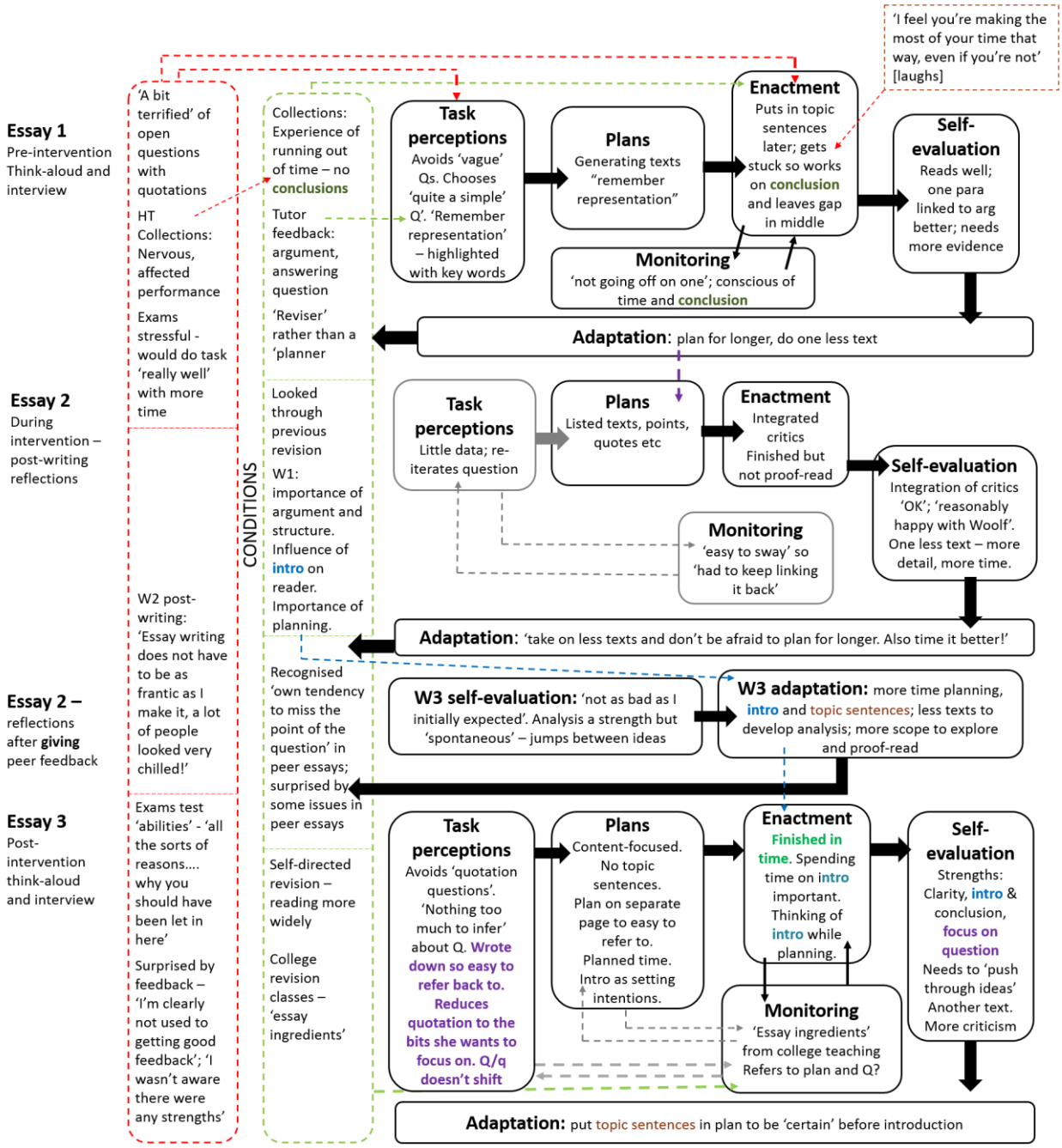
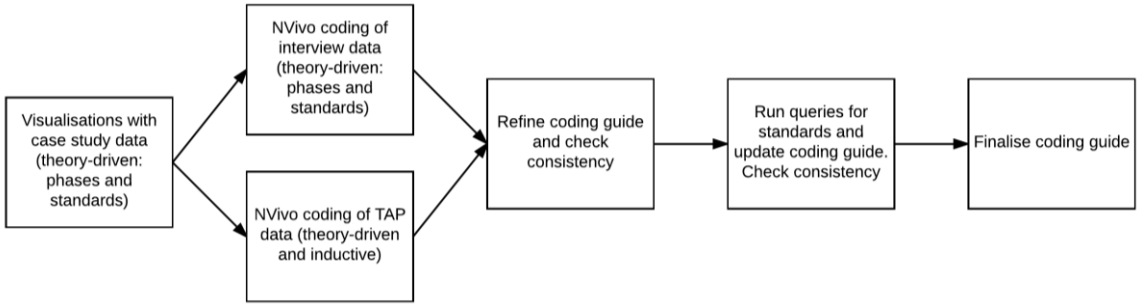


Figure 6: Digitised visualisation (Chloe)

Figure 7: Development of SRL coding guide



The guide uses the same codes as for the visualisations, that is the phases of SRL in the Winne and Hadwin model, plus further codes that were developed during coding to capture the specific process of this context. There are broadly two types of code: phase and cross-phase. As noted in the guide, some of the coding was initially done through NVivo's automated process as particular sections of the interviews were designed address particular phases of learning. The first round of manual coding comprised of thoroughly checking and amending these as necessary, and applying the 'standards' code. For each case, I then checked the output of the standards code as in the initial stages it was easy to confuse standards with beliefs about quality in writing – data was only coded against standards when these were used to make a judgement on writing.

To trace which standards were used in particular phases of learning, queries were run for each phase, retrieving text coded at each phase AND standards. The results were printed with coding stripes, and I hand checked each document, asking for each item whether it was really an example of a standard, and whether the criteria coding adequately captured the substance of the standard. New codes were created at this point for sub-standards. Queries were also run for each phase and NOT standards to check for any data which should be coded as a standard. As SRL is recursive, some data is coded at multiple phases – for example, a participant may report their planning, and make a self-evaluation on their planning in the same utterance. I tried to see if it was possible to

separate such overlaps, but it was very complex, and ultimately not essential to the research questions.

As the think-aloud protocols were more fine-grained, further data-driven codes were developed to capture students' operations during the task perceptions and planning phases. Theory-driven coding was also partly used to address the research question on feedback. As described in the next section, a coding guide from Patchan et al.'s (2016) feedback model was adapted.

4.11.5 Analytical strategies for each research question

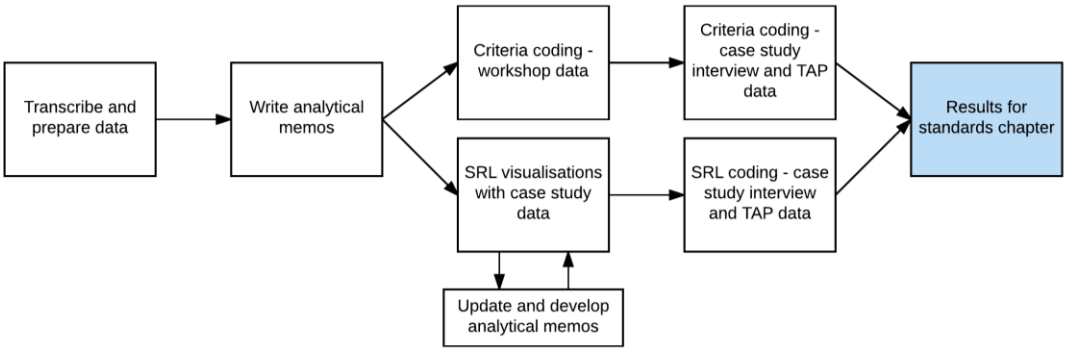
An overview of analytical strategies for each research question is given in Table 8; this is now briefly elaborated on, with more detailed descriptions given in the findings chapters.

RQ1. During the workshops, which criteria do students use to evaluate written essay products?

The analysis for this research question is at the group level. The main analytical resource for this research question is the criteria coding (data-driven, NVivo). As described in Section 5.2, coding queries were used to compile tables to summarise the frequency of references. To create visualisations to show how criteria are related, a matrix coding query was run in NVivo to identify where references were coded at pairs of criteria. The visualisations were initially hand-drawn, then digitised.

Table 8: Overview of Research Questions and Analytical methods						
	Over-arching: development of self-regulation of writing	RQ1. Criteria	RQ2. Standards	RQ3. Task perceptions	RQ4. Feedback	RQ5. Self-efficacy
Analytical memos (by hand, data-driven)	X	X	X	X	X	X
Criteria coding (NVivo, data-driven)	X	X	X	X	X	
Visualisations (by hand then transcribed; theory-driven)	X	X	X	X	X	X
SRL coding (NVivo, theory-driven)	X		X	X	X	X
Feedback coding (NVivo, theory and data-driven)					X	
Statistical analysis of self-efficacy questionnaire						X

Figure 8: Analytical process for RQ2



RQ2. Which criteria do students use as standards to self-regulate their essay writing processes?

The analytical process for this research question is shown above in Figure 8 above. Chapter 6 draws on criteria coding, SRL coding and visualisations to present both an overview at group level of which criteria became standards. It also presents in-depth analysis of one case study writer's use of standards.

RQ3. (How) do students' task perceptions shift over the course of the study?

Analysis of task perceptions is primarily case-by-case so the key strategy was visualisation, supported by theory-driven SRL coding and analytical memos. Chapter 7 also draws on coding queries which identify where the *question* and other related codes are used over the phases of self-regulated writing.

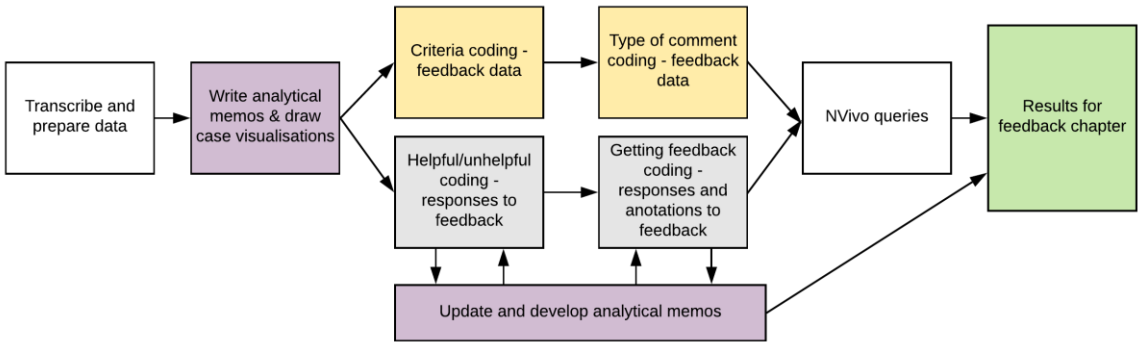
The analytical approach for this RQ has shifted during the study. While visiting University of Victoria, I attempted to develop a coding guide from the Hadwin and Winne (2012) layers of task perceptions: explicit, implicit, socio-contextual. Through readings and annotations of interviews, think-aloud and workshop data for five cases, I developed operationalisations for these layers appropriate to the Oxford context. However, discussion with lab members in Canada threw up differences in interpretations of the layers, and after coding interviews from the five cases, I found that the codes were not good fit for my data. In contrast to Canada, UK students specialise in a discipline from the beginning of the degree. Therefore, it was very hard to judge where perceptions were at the implicit level or the socio-contextual level. In addition, Oxford examination papers give very few instructions, so the explicit code was very low frequency. A further confounding factor was that participants chose their questions, and so there was a great

variety of questions and so task perceptions. Initial categorisations of data using the three layers of task perceptions are given in Appendix K. While the layers of task perceptions remain a useful analytical tool, I decided that a case-by-case approach is more productive for answering the research question. The decision to shift approach was an important reflexive moment. For me, coding with NVivo had become the default analytical approach as I was already using it for criteria coding. It is also widely used in the literature, and it was important to challenge this orthodoxy.

RQ4. (How) does peer feedback contribute to students' self-regulation of writing processes?

This analysis draws on both case and group level analysis, and Figure 9 summarises the process. At case level (in purple), visualisations and analytical memos are used to identify whether and how case study participants use feedback, and to link responses to feedback to recipients' beliefs and standards for writing.

Figure 9: Analytical process for Research Question 4



At group level (in yellow), the reviewers' feedback comments and annotations were coded using the criteria guide, and in addition, a 'type of comment' guide. For the latter, four codes from Patchan et al. (2016) were used as starting point. They were: praise, problem, solution and localisation. Further codes were added to capture the nuances of these particular data, as explained further in Section 8.3. All feedback coding guides can be found in Appendix L.

In addition, at group level, I coded participants' responses to and perceptions of feedback (in grey). Two coding guide were used. Firstly, in preparation for a paper presentation at SRHE in December 2015, I developed inductively a coding guide to capture students' perceptions of helpfulness and unhelpfulness. Later, in 2017, codes were developed to track participants' receptiveness to feedback, in a shorter coding guide named 'getting feedback'. This was applied to annotations made on received feedback, and Workshop 3 reflections. The 'type of comment' and 'getting feedback' guides were shorter and less technical, so the coding process with Nadri, the practical support assistant, was more dialogic. For 'type of comment', we often discussed the difference between problems and solutions as the distinction in these data was quite nuanced.

RQ5. What is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing?

This research question is addressed using the self-efficacy questionnaire data. To test the null hypothesis that is that there is no difference in students' ratings of self-efficacy at Workshop 1 and at Workshop 4, I ran a Wilcoxon signed-rank test to compare students' ratings of self-efficacy on each item. The results are given in Section 5.9, with further analysis in Appendix C. The quantitative results are contextualised with case study data, drawing on case study analysis,

particularly SRL coding. Chapter 7 examines self-efficacy in the context of task perceptions, and Chapter 8, feedback.

Over-arching research question: (how) does students' self-regulation of writing develop in the context of peer assessment writing workshops?

The five research questions feed into the overarching research question. However, the theory-driven SRL coding, analytical memos and visualisations are of particular value for tracing participants' development over the course of the study. Table 12 in Section 5.6 is a summary table, tracing students' large-scale adaptations over the course of the study. More detail on its production is given there, and Appendix H gives an extract from an analytical memo as an example.

4.11.6 Analysis of essays

As the focus of the thesis is on writing processes, essays – writing products – are used to complement the interview, think-alouds and workshop data, rather than being a key source. The emphasis of this study is on intentionality rather than automatic or unconscious behaviours. Therefore, the essays themselves are used to trace evidence of strategies that participants themselves identify in workshop data and mention in interviews.

Revisions could be seen as a trace of self-regulation during writing. The frequency of revisions could also hint as to whether writers frequently self-monitor and revise writing accordingly, or alternatively the ease with which they write about the particular topic. Revisions are evident in the handwritten essay data as crossed out words, inserted words, and instructions to read a sentence as the start of a new paragraph. However, I did not use discourse interviews specifically targeting

revisions or video data, which would be essential to tracing intentionality. Within the writing research community, revisions are not always interpreted as meaningful, especially without additional data. To borrow the terms used at one research meeting at the 2016 EARLI SIG Writing conference, such revisions could be ‘an act of meaning-making’ or ‘mere noise’ (Lindgren et al., 2016). Thus, revisions are not included in the analysis unless there is a specific mention of revisions in interview or think-aloud data. This decision also aligns with the study’s main focus on self-regulation over time, rather than at a micro level.

The impact of the intervention on writing quality may be investigated after doctoral study. This is a complex research question in itself as to align with the overarching research question, quality would have to be defined developmentally. An ecologically valid way of measuring writing quality would be to have raters use the course assessment criteria, as most intervention studies in 'naturalistic' settings seem to do (e.g. Wilder and Wolfe, 2009) unless a rubric is provided for student use (e.g. Cho and MacArthur, 2011). However, according to examiners' reports, 'assessors tend to ‘bunch’ marks, shying clear of giving high Firsts or low 2.2s – a problem raised every year by external examiners' (Examiners' report, 2013, p.26). This means that there is unlikely to be a normal distribution, with few marks in the 50s and below and in the high 70s and above. Such a distribution would make it difficult to capture any progress as writing development is very incremental and slow (Camp, 2012; Alamargot and Fayol, 2009). Publicly examined work is double marked, with a third examiner brought in if there is a discrepancy of a degree class or more. This is quite a high tolerance and implies that the faculty accepts there is disagreement between markers, and no easy consensus on quality. So, an alternative way of looking at writing quality would be needed.

Creating an alternative rating scale would be time-consuming and not necessarily reflective of the values of the faculty. As Hughes (2011) argues, an ipsative approach is better suited to assessing students' progress than criterion-referencing. One possible method would be to use adaptive comparative judgement, where judges are asked to compare two pieces and simply judge which is stronger. It is claimed that this offers reliable ratings without assessment criteria (Pollitt, 2012). However, it seems that this has not been used ipsatively as yet, or to trace development over time. Thus, writing quality is an interesting research question but a challenge beyond the scope of the thesis.

4.12 Limitations of the research design and methods

An obvious limitation of the study is that it draws on a relatively small sample of participants. However, the dataset is rich, detailed and extensive, at approximately 250,000 words. The study does not aim to generalise, but to examine the promise of peer assessment, so this is adequate for the purposes. In addition, the research design does not allow for any gains in achievement or writing quality to be measured, neither as a single-group design nor against a comparison group. Achievement is of great interest to researchers, practitioners and policy makers, and this could limit the impact of the study. Nonetheless, a fine-grained focus on learning processes should contribute to understanding about how peer assessment works and inform the design of better measures in future research.

Due to my commitment to naturalistic research that would hopefully benefit participants' learning, several choices were made that eased the research burden but could have enriched the research data. For instance, using discourse-based interviews (Odell, Goswami and Herrington, 1983) or video-recording writing and using stimulated recall could have provided richer insights

into participants' decisions and processes. Equally, while the additions participants made during the workshops offer some traces of what they took from discussion, it would have been ideal to record the teaching sessions. This would allow for individual learning to be situated within the social situation.

The intervention was run as a standalone course. This is a strength in research terms as it allowed fidelity and consistency of implementation, and all who participated were volunteers, which was important given. However, in future iterations of this design-based research, the workshops should be embedded into the degree course to examine further how it interacts with other teaching, and its feasibility. This issue is examined in further detail in the discussion chapter.

Anonymising the essays and feedback is also time intensive for the teacher/researcher. However, much of the labour intensive work stemmed from the essays being written by hand, as they were to be in the exam. Oxford is a very traditional setting, and in other situations where work is typed, it would be much easier to conduct peer assessment online. This study is in some sense an anomaly in the wider literature as much peer assessment research through platforms such as MyReviewers (Moxley, 2013) and SWoRD (Schunn, 2016).

4.13 Chapter conclusion

This chapter has detailed the research design and the methodological decisions made. The chapter describes how I strived to balance the need for high quality data while managing the burden on the participants. The dataset of approximately 250,000 words covers a range of data, allowing self-report to be contextualised and challenged with traces. A further strength of the study is the combination of analytical strategies, which together give a holistic and fine-grained perspective on

participants' self-regulated writing over time. The next chapter presents headline findings and concepts, focussing on the criteria identified in students' comment on writing, the large-scale adaptations students made to their writing processes, and the impact of the workshops on self-efficacy. Taken together, these findings set up the in-depth discussion in later chapters.

Chapter 5

What students notice about writing and choose to adapt

> 'SRL is viewed as *proactive* processes that students use to acquire academic skill, such as setting goals, selecting and deploying strategies, and self-monitoring one's effectiveness, rather than as a reactive event that happens to students due to impersonal forces' (Zimmerman, 2008, pp.166–167; original emphasis).

5.1 Introduction

Zimmerman's definition of self-regulated learning is a worthwhile opener to the findings chapters. We begin with Holly, who is particularly proactive, and yet whose experiences raise key issues addressed in this chapter and the wider study. Holly actively sought and created opportunities to develop her learning and writing. When we met in February, she explained she was keen to take part as the workshops would involve assessing example essays, 'something we've all been begging our tutors for'. She and others had been struggling to understand the conventions of commentaries, and tutors had agreed to share model answers. In addition, to prepare for her January collections – college formative exams – Holly reported that she had continued with a strategy she started using at school – writing timed practice essays, then after a short break, 'marking' them. She deconstructed each essay line by line, and then came up with an overall judgement. However, although Holly was very aware that she needed to move beyond A-level, she reported that she didn't have enough understanding to really 'self-monitor [her] effectiveness', to quote Zimmerman. She felt unable to use the marking criteria provided by the English Faculty to self-assess:

> 'I think, because I find the Prelims criteria quite vague, I can't really use them to mark my own writing. I just go… I think I base it more on, um, from my A2 knowledge, does it meet and excel [sic] that? Because if it just meets, then that's obviously not good enough, but also does it meet my own personal

> standards? Like, would I be happy to have written this in a Prelims exam? […] But, yeah, it's, it's - the Prelims criteria, it said, 'Writing eloquently', for example. I went, "Hmm, OK…" Well, I haven't got a standard of eloquence to mark myself against. I feel fairly eloquent so, unless someone tells me otherwise, that's fine. But… yeah. The Prelims criteria are not something that I really use, which I think could be a problem because, obviously, if you're being marked against the Prelims criteria you should normally be trying to meet them, but there's definitely that sense of kind of, it's not even technical jargon that I don't understand – it's just not having a benchmark.'

Holly's reflections exemplify the obstacles that self-regulating learners face. There is a real limit to the extent that students can be proactive when what they are aiming for is unclear. Holly can keep self-assessing her work to '[her] own personal standards', but she is aware that it could be problematic if these do not align with what tutors and examiners value. In an examination situation, her considerable efforts may not be recognised by markers. Holly's reasoning also resonates strongly with Sadler's (2010, p.545) arguments about the limitations of 'telling' students about quality through feedback and assessment criteria. The 'technical jargon' is not the issue, but rather the lack of a 'benchmark', or the implications and tacit knowledge behind terms such as 'eloquence', and the meaning for her own writing. Holly's requests for examples and participation in the workshops shows she is aware that she needs to 'see' quality, as Sadler (2013, p. 8) puts it, to really develop her writing.

This chapter examines what students 'see' in example essays, and how this both informs the adaptations they make to their own writing, and their self-efficacy beliefs. By presenting a broad overview of headline findings and concepts, the chapter sets up the threads that weave through the later, more detailed chapters on the key areas of task perceptions, standards and feedback. The chapter begins with an overview of the criteria that run through students' evaluative talk, comment and feedback on writing. These were identified through inductive coding, and the results are visualised to show the complexity of writing quality, and how criteria interconnect. Next, the

criteria codes are compared with the Faculty criteria to which Holly refers, and the lists workshop groups formulated after assessing example essays. There is considerable overlap between the three sets of criteria, which contributes to a validity argument for the coding, and for assessing example essays without pre-determined criteria. While the Faculty and student criteria focus exclusively on declarative qualities of writing, the coding shows that students also comment on procedural aspects of writing. These procedural criteria codes play an important role in the adaptations that students make to their writing, which is the focus of the next section. The challenges of exam writing prompt writers to make adaptations to often long-held approaches. I give examples of common strategic changes and whether and how they were motivated by the workshops. Next, I discuss barriers to adapting writing, and two outlier cases where students are satisfied with their approach and make no further plans to adapt. Self-evaluations and self-efficacy are important to adaptation decisions, and these findings link into the results of the self-efficacy questionnaire. In this section, I give an overview of the results, connecting quantitative results to qualitative insights about the impact of the workshops on self-efficacy beliefs.

5.2 Identifying criteria in students' talk, comment and feedback on writing

This section addresses RQ1: During the workshops, which criteria do students use to evaluate written essay products? It gives an overview of the criteria codes identified in students' evaluative talk, comments and feedback and visualisations to show how they are interconnected. Unless otherwise stated, the term 'criteria' is used in the findings chapters to refer to the criteria codes identified in participants' discourse about writing across the dataset, rather than the Faculty or student groups' criteria.

To answer the research question, the data were coded in NVivo 10, as described in Section 4.11, and recapped briefly here. The coding guide is included in Appendix I. During the workshops the criteria were not pre-determined but arose from holistic assessment and discussion of example essays, so the coding was conducted inductively. No coding process can be truly free of external influences and biases, but after looking at the Faculty criteria to design the workshops in late 2014 to early 2015, I did not return to them until late 2017, only once all analysis, including the visualisations, had been completed. As described in 4.11, the starting point was responses and annotations to example essays: traces of evaluative thought and talk during the workshops. To preserve meaning, data was coded as idea units rather than sentence clauses – participants often elaborated on statements, so sometimes an idea continued between clause or sentence boundaries. It soon became clear that a hierarchical structure would be inappropriate as very different criteria often co-occurred. Thus, idea units could be coded at multiple criteria as it was often impossible to disentangle them. Through multiple iterations, the coding guide was refined so that each code was distinct. In the case of two similar but distinct codes, for example, *depth and detail* and *development*, coding had to be mutually exclusive, with the former referring to material and analysis, and the latter to argument, and the extent to which a writer followed through with a point. As a result, the coding guide has a flat structure, with no sub-codes. The final refined version applied across the dataset contains thirty-four distinct criteria codes. The coding guide only captures the content or idea behind each criterion, not the levels of achievement within it; the emphasis in coding was capturing the underlying topic or noun, rather than descriptive adjectives, e.g. good/solid/poor. Nonetheless, a separate coding guide was developed to identify whether feedback data conveyed praise, a problem or a solution. These results are discussed in Chapter 8.

Table 9: Overview of criteria coding

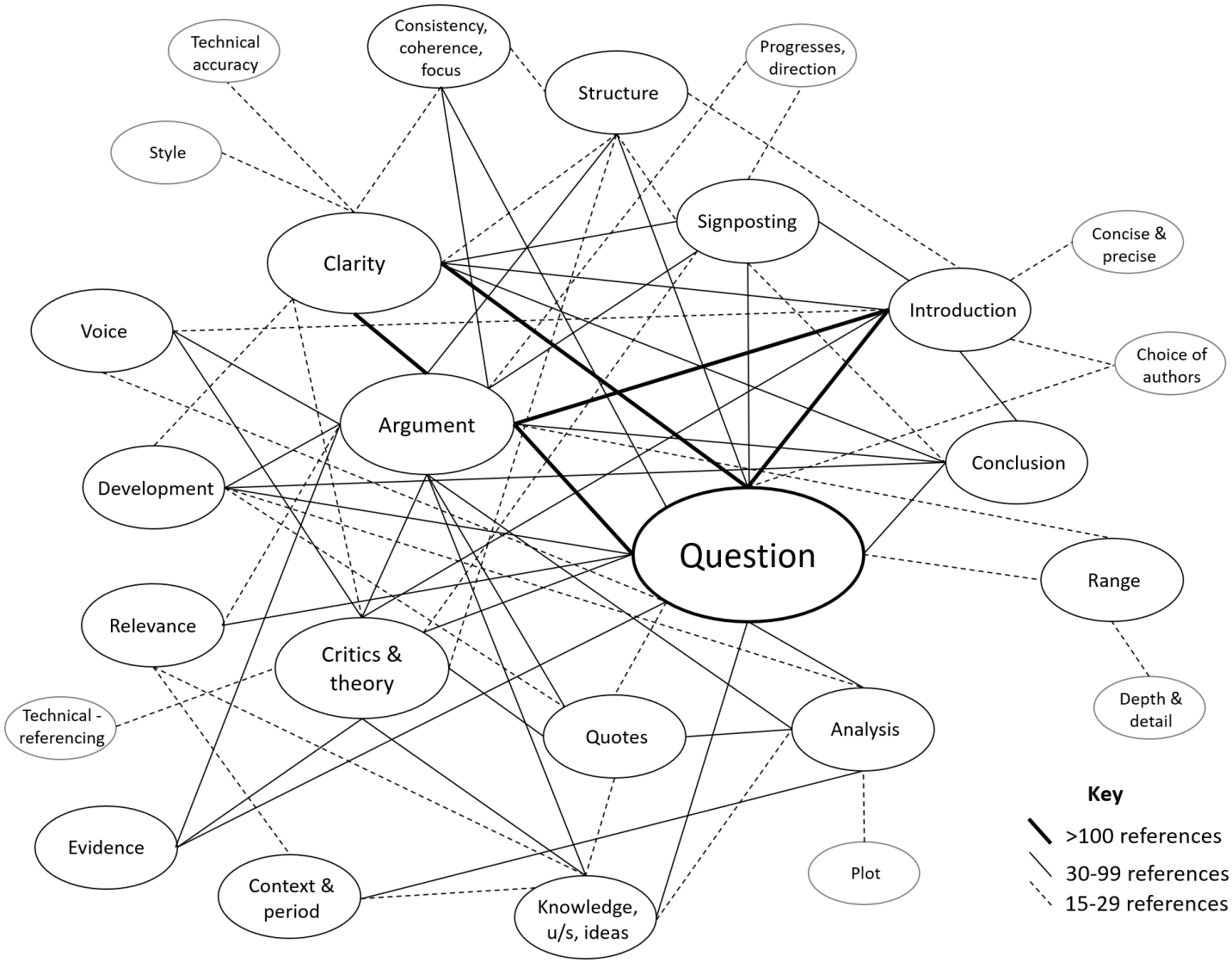
Asterisks (*) denote the presence of the code in Faculty and/or student criteria, discussed in 5.3 and 5.4 below.

Criteria code	References	%	Identified in Faculty criteria	Identified in student group criteria
Question (C21)	1170	15.2%	*	*
Argument (C2)	674	8.7%	*	*
Critics and theory (C9)	627	8.1%	*	*
Clarity (C4)	536	7.0%	*	
Introduction (C15)	407	5.3%		
Signposting, topic sentences & paragraphing (C26)	355	4.6%		*
Conclusion (C6)	308	4.0%		
Quotes (C22)	288	3.7%		
Analysis (C1)	266	3.4%	*	*
Development (C11)	248	3.2%	*	
Context and period (C8)	232	3.0%	*	*
Knowledge, understanding and ideas (C16)	231	3.0%	*	
Evidence (C12)	214	2.8%	*	*
Technical accuracy (C31)	208	2.7%	*	
Style (C28)	196	2.5%	*	
Structure (C27)	191	2.5%	*	*
Voice, positioning of the author and opinion (C33)	184	2.4%		*
Range; scope (C23)	168	2.2%		*
Consistency, coherence and focus (C7)	159	2.1%	*	*
Comparison and linking (C5)	146	1.9%		
Relevance (C24)	138	1.8%	*	*
Choice of authors or texts (C3)	92	1.2%		
Depth and detail (C10)	92	1.2%	*	*
Technical – referencing (C30)	78	1.0%		
Style - concise and precise (C29)	75	1.0%		
Nuance; complicating (C17)	72	0.9%	*	
General; broad (C14)	68	0.9%		
Word count and essay length (C34)	54	0.7%		
Progresses; direction (C20)	48	0.6%	*	*
Plot (C19)	47	0.6%		*
Specific knowledge (C25)	42	0.5%	*	*
Original (C18)	36	0.5%	*	*
Terminology (C32)	35	0.5%		
Finishing (C13)	26	0.3%		
Total	**7711**	**100%**		

Table 9 gives the total number of references for each distinct criteria code across the self-report and trace dataset: think-aloud protocols, interviews, annotations, additions and responses to example essays, annotations and feedback on peer essays, and reflections written at the end of each workshop. The table also gives the percentage of total references attributable to each code. *Question* is the most frequent code by quite some distance, followed by *argument*. The contribution of *question* to writing quality is significant for RQ3: '(How) do students' task perceptions shift over the course of the study?' These findings are discussed in Chapter 5. An asterisk in the final two columns denotes whether a code can be identified in the English Faculty criteria, and those formulated by student groups during the workshops. The majority of the most frequent criteria codes also feature in the Faculty and student group sets, indicating a good overall fit. These relationships are discussed in more depth below.

As the coding structure was flat rather than hierarchical, examining the relationship between codes offers insight into how different criteria inter-relate to form an underlying concept of quality in exam writing. Figures 10 and 11 are visualisations of the criteria, and represent how codes intersect, or the frequency of references at two criteria. Figure 11 shows only a subset of the most frequent criteria. To create these visualisations, a matrix coding query was run in NVivo to identify where references were coded at pairs of criteria. The resulting cross-tabulation can be found in Appendix M. The cross-tabulation was used to draw the visualisation, first by hand and then digitally. As the key shows, the thickness of the connecting lines denotes the frequency of co-occurrence. The visualisation includes only instances where there were 15 or more co-occurrences, and so does not exhaustively capture each and every co-occurrence; some lower frequency criteria are therefore not included. To give an example of how to read the visualisation, *analysis* is supported by *quotes*, and can be supplemented by information about the *context and*

Figure 10: Visualisation of criteria coding



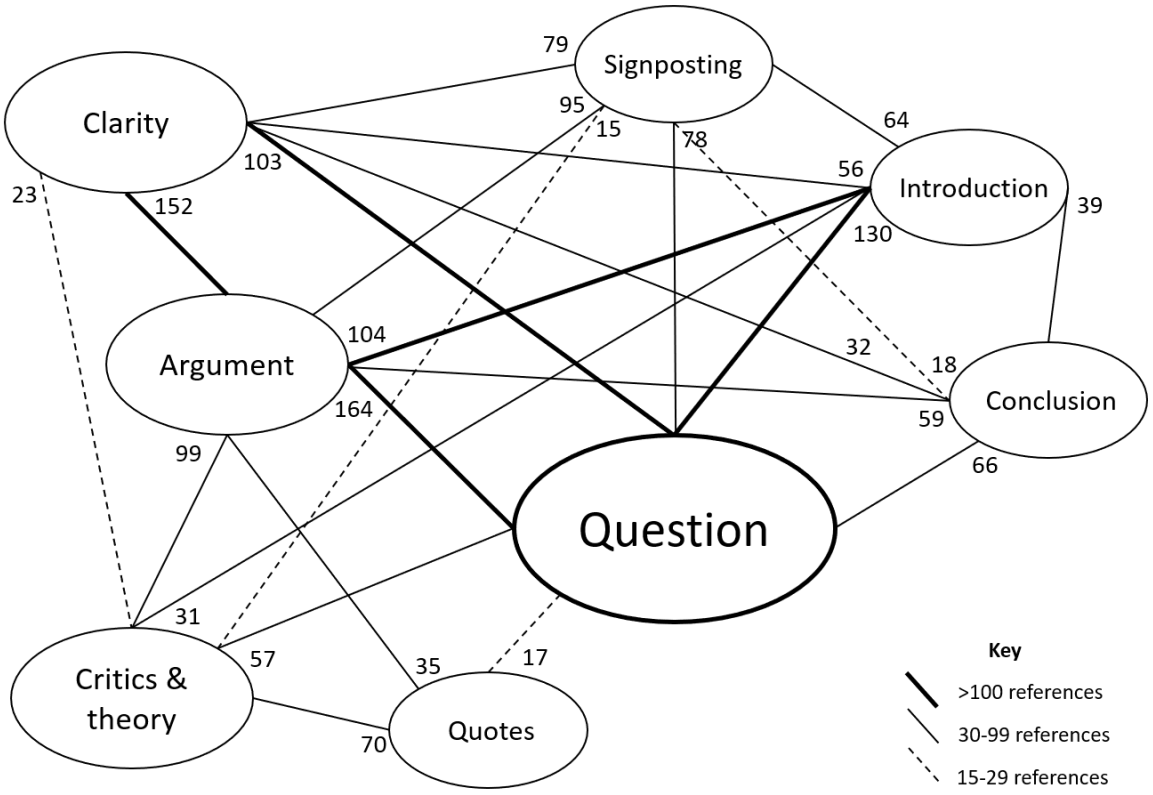


Figure 11: Visualisation: subset of the most frequent criteria

period. It is more weakly linked to *plot*, which in the wider data, is mostly considered to undermine rather than strengthen *analysis*. More broadly, *analysis* is linked to and supports the essay's *argument* and response to the *question*.

The literature review chapter highlighted how terms used to describe and evaluate academic writing are mutable, and slippery in meaning (e.g. Lea and Street, 1998). The value of the visualisations in particular is that they show how, according to students' evaluative comments, these concepts interconnect in exam writing, and which are most frequent and so implied to be most important. The centrality of *question* to writing quality is immediately apparent. Not only is it the most frequent code used in evaluative comment on writing, but it is at the heart of the visualisations, with all but a very few low frequency criteria co-occurring with it. Addressing the

question is not only a matter of task compliance (Sadler, 2010), but is also at the centre of writing quality more broadly. The subset of criteria visualised in Figure 9 shows how *question* is particularly strongly connected with other high frequency criteria: *argument* (164 references), *introduction* (130 references) and *clarity* (103 references). The following extracts from workshop data show how responding to the question strengthens the quality of a writer's argument, and how other writing qualities are also involved or related. The extracts were retrieved from a query for data coded at *argument* AND *question*. In the first comment, 'colonialism' refers to the question. In the third comment, 'quote' refers to the quotation included in the question.

> 'The essay on Forster and Woolf was the strongest as it followed a clear line of argument that simultaneously nuanced and challenged the binary implied by colonialism' (Joe, reflections on example essays, Workshop 1; also coded as *clarity*, and *nuance; complication*).
>
> 'This essay does not seem to have left itself time to develop an argument, let alone one which would address the terms of the question and the complexity of the quote which was given.' (Abbie, response to an example essay, Workshop 1)
>
> 'I tried to make my argument clear from my introduction. There was a point when I noticed the final sentence of the quote 'remould our life' & I felt I had to respond.' (Jess, reflections on her own essay, Workshop 2; also coded as *clarity* and *introduction*).
>
> 'In places, points were far more implicit than explicit, particularly in their relevance to the question – I could see the argument was there but it just wasn't clearly linked – and I know I do this in my own writing.' (Celine, reflections on peer and own writing, Workshop 3; also coded as *clarity* and *relevance*).

The extracts highlight the importance of connecting *question* and *argument*. They also contain possible strategies that writers can use to achieve this: nuancing and challenging underlying ideas, addressing terms, noticing and responding to a particular part of the quotation, signalling the relevance of points to the question. *Clarity* also recurs throughout these extracts; it is important, for example, to make the argument clear from the introduction. The importance of *clarity* is

particularly prominent in Chapter 7, which examines how participants' metacognitive standards for writing shifted over the course of the study, and how students, as Jess hints at here, became more focussed on the reader's experience. The connections mapped in the visualisations also show that beyond lofty, broad criteria such as *clarity,* some criteria are more procedural, or strategies for achieving other criteria. The visualisation summarises how criteria such as *clarity* are reached: through *signposting*, through the *introduction*, and to a lesser extent, through *style*, a high level of *technical accuracy* (e.g. spelling), and *consistency* - maintaining *coherence* and *focus* of argument throughout the essay.

The visualisations represent the complex, tangled web of criteria that underlies writing quality, itself a unitary construct (Sadler, 2010). The criteria codes together present different combinations, different possibilities of achieving writing quality. However, it should be noted that not all of these criteria are relevant to all essays. Some criteria are marginal, such as making and communicating a fitting *choice of authors*. The *choice of authors* may be pertinent only if, for example, the authors chosen for an answer are particularly interesting, or alternatively seemingly irrelevant and unexplained. A small number of criteria, while relatively frequent, are not featured on the visualisation as they do not often co-occur with others. I would contend that *question* is always relevant, as it is central to the web and to task compliance, to the essence of examination writing. To use Sadler's (1985) terms, it is a manifest criterion, whereas *choice of authors* may often be latent. Knowing which criteria are appropriate to refer to and when is central to professional practice in assessment, as sometimes latent criteria can make or break quality (Sadler, 1985; Wyatt-Smith and Klenowski, 2013). I argue that this is also important for self-regulating writers: noticing which criteria contribute most to quality, and using this knowledge to decide which adaptations to make, as discussed below in Section 5.6. The discussion now considers the relationship between the criteria coding and two sets of criteria to which student

writers had access: the official English Faculty marking criteria, and the sets of criteria workshop groups formulated after assessing examples in Workshop 1 and Workshop 2.

5.3 How the coding relates to English Faculty criteria

While the coding guide captures writing quality according to student writers and peer assessors, the English Faculty criteria are written and used by disciplinary expert readers. The Faculty's 'marking and classification criteria' for Prelims, Finals and college collections are shared with students at the beginning of their degree in the course handbook. Table 10 gives the 'criteria for examination questions'; criteria for portfolios are slightly different. The seventeen criteria are presented in four groups: engagement, argument, information, organisation and presentation. The handbook also includes three pages of mark descriptors, outlining how each range or band of numerical marks corresponds to a category (distinction/pass/fail) and to the criteria. Descriptions of performance are given, but particular criteria are not specified. Candidates are not necessarily expected to reach the same level of performance on all criteria groupings. Although there is no description of how marks are allocated, assessment seems to be holistic, with the possibility of strengths and weaknesses balancing each other out. For example, the description for 70-74 – Distinction – states:

> 'Scripts will be at least very highly competent across the board, and probably excel in at least one group of criteria. Relative weaknesses in some areas may be compensated by conspicuous strengths in others.'

The mark descriptors document begins with the highest band: '86+' – distinction. This descriptor makes no reference to criteria: 'scripts will be so outstanding that they could not be better within the framework of a three-hour exam. Work of marked originality and sophistication'. At this level, the criteria become irrelevant. At other levels within Distinction, work 'excels'; pass level work is

described in variations of 'competence'. Towards the lower levels of a pass, a deficit discourse is used: essays 'fall down on one or more criteria', or 'will be marred by a failure on one criterion or another' (50-59).

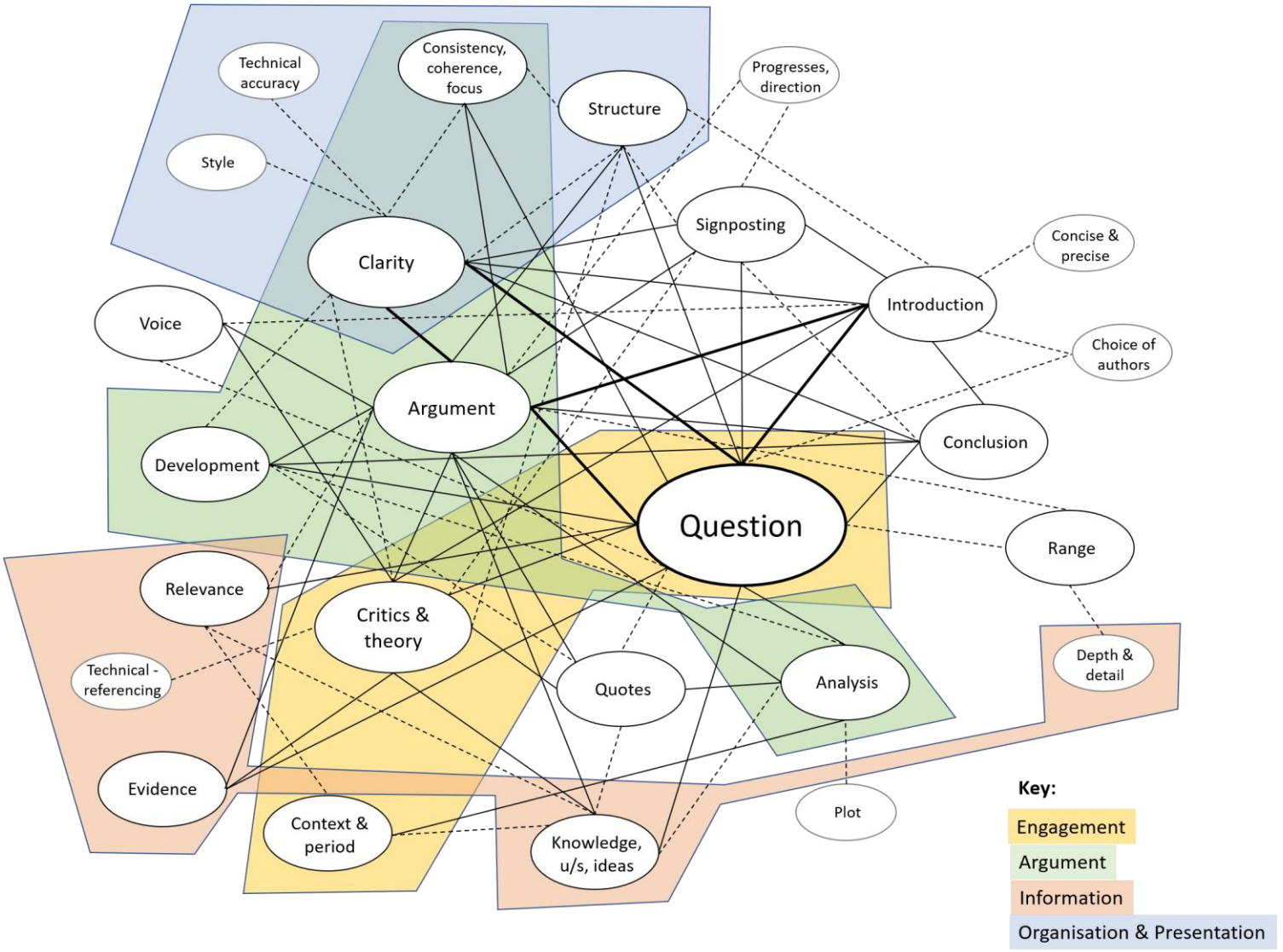
The Faculty criteria are presented in Table 10. For ease of reference in the next few pages, each is assigned a label (Eng1, Eng2 etc). The final column identifies the equivalent codes from the criteria coding guide. Almost all Faculty criteria can be linked to the coding guide. The only criterion not to map directly onto the coding guide is 'accuracy of facts' (Inf3); this is unsurprising given that the workshops did not focus on subject content. Similarly, 'intellectual penetration and sophistication of conceptualisation' (Arg3) could relate to *nuance; complicating* and *development,* but is otherwise beyond the scope of the coding. Both criteria very much require a disciplinary expert's judgement, and so are perhaps less relevant for first years students deciding how next to develop their writing.

The other criteria, however, can all be linked to the coding guide. Clusters can be identified in the visualisation of criteria codes, as the web of connections necessitates that similar but distinct codes clump together. Figure 12 shows how the four groups of the Faculty criteria can be mapped onto the visualisation. The figure uses background shading to show map each group onto the visualisation: engagement (yellow), argument (green), information (orange), organisation and presentation (green). Codes which are not directly identifiable in the Faculty criteria are not shaded. This visualisation shows that the aspects of writing that students comment on – the codes – are in overall alignment with faculty criteria. The importance of *question* is reflected both in the centrality and frequency of the code in the visualisation, and its prominence in the Faculty criteria,

Table 10: Mapping English Faculty criteria onto the coding guide

English Faculty Criteria		**Label**	**Coding guide equivalent**
Engagement	- incisiveness of engagement with the question;	Eng1	*Question*
	- depth and sophistication of comprehension of issues and implications of the question;	Eng2	*Question*
	- relevant awareness of literary history and theory and critical traditions;	Eng3	*Context and period; critics and theory*
	- directness of answer to the question;	Eng4	*Question*
	- grasp and handling of critical materials.	Eng5	*Critics and theory*
Argument	- coherence of argument;	Arg1	*Argument; consistency, coherence and focus; progresses, direction*
	- analytical clarity and power;	Arg2	*Analysis; argument; clarity*
	- intellectual penetration and sophistication of conceptualization;	Arg3	*Nuance, complicating Development*
	- originality of argument;	Arg4	*Original; argument*
	- quality of critical analysis of text in the service of argument.	Arg5	*Analysis; argument*
Information	- relevance of deployment of information;	Inf1	*Relevance*
	- depth, precision and detail of evidence cited;	Inf2	*Evidence; depth and detail; (quotes; technical – referencing)*
	- accuracy of facts;	Inf3	? Beyond the scope of the workshops
	- relevant knowledge of primary texts.	Inf4	*Knowledge, understanding and ideas; relevance*
Organisation & Presentation	- clarity and coherence of structure;	OP1	*Clarity; consistency, coherence, focus; structure*
	- clarity, fluency and elegance of prose;	OP2	*Clarity; style*

Figure 12: Visualisation showing how Faculty criteria map onto coding



with three of the five 'engagement' criteria directly related. The visualisation also highlights that while criteria can be clustered, relationships cut across these groupings, and different facets of writing are inter-related. In the Faculty criteria, the key words, 'clarity' and 'coherence' appear in both the 'argument' and organisation and presentation groups (Arg1, Arg2, OP1, OP2).

Many high frequency codes are left unshaded: *signposting, introduction, conclusion, quotes, voice, range*. These codes are not unimportant in the student data. While the Faculty criteria describe ideal written products, these unshaded codes reflect procedural aspects of writing. The visualisation shows how these criteria codes link to and contribute to Faculty criteria groups. For example, *signposting* contributes to all four Faculty criteria groups, and in particular, the prominent codes *clarity*, *argument*, and *question*. The interconnectedness of *signposting* indicates its importance from a student writer's perspective. Such unshaded codes represent many of the aspects of writing that students choose to change and adapt in order to meet loftier, broader criteria which are given more prominence in the Faculty's descriptions. This is discussed in further detail later.

5.4 Student group criteria

The discussion now turns to the criteria that student groups formulated during the workshops, and their relationship both to the codes and to the Faculty criteria. At the end of Workshop 1, after assessing and discussing three example essays, the following questions were displayed to foster discussion:

> 'Which aspects of writing have been important in our discussion of the example essays? If we were to treat these aspects of writing as criteria, what would we call them?'

Through discussion, each group produced a list of criteria, which was written up on the whiteboard. These were emailed to participants after the session, then reviewed and sometimes extended in Workshop 2, between assessing the fourth example essay and writing their own. It was emphasised that participants were not limited to these criteria when making judgements or writing their own essay; rather they were used to summarise discussion. The criteria were displayed as a reference point in Workshops 2, 3 and 4, and the strategies handout students received in Workshop 4 contained a synthesis of all groups' criteria, which itself was discussed.

Table 11 presents the workshop groups' criteria, with columns to identify their equivalents in Faculty criteria and the coding. The students' criteria are more realistic and procedural than the Faculty criteria. For instance, Groups 1 and 3 both mention *signposting*. The criteria also include quite specific instructions, for example, how to write about texts in appropriate detail: 'balance between one-off and close reading of episodes without too much plot' (Group 2). Two student groups include 'confident tone' and 'confidence' (Groups 1 and 3). Group 2 also refer to 'creating and sustaining your own argument'. These map onto the code *voice,* which is left implicit in Faculty criteria, only hinted at in 'analytical clarity and power' (A2) and 'clarity, fluency and elegancy of prose' (OP2). It seems that assessing and discussing example essays raises the importance of 'confidence' and having 'your own argument', a more realistic way of achieving power in essay writing. The workshop group criteria also reflect student interests and idiosyncrasies. For example, the final criterion for Group 3 reflects how students from one college were beginning to see connections between papers as they began to prepare for the Paper 1 portfolio in Weeks 4 and 5.

As well as including procedural, realistic strategies, each workshop cohort covers the four Faculty groups, aside from Group 2, who don't directly address Organisation and Presentation. Key words

Table 11. Mapping workshop groups' criteria onto the English Faculty criteria and coding guide

Key: Blue indicates a keyword from the faculty criteria groups (engagement, argument, information, organisation and presentation). In the Faculty labels column, **bold** indicates similar wording to the Faculty criterion. Brackets indicate a link requiring more inference.

	Faculty criteria	**Coding guide**
Group 1		
Addressing the quotation / question	Eng4 (Eng1; Eng2; Eng5)	*Question*
Formulating an argument in response	Eng4	*Argument; question*
Setting up the structure (e.g. topic sentences)	OP1	*Structure; signposting…*
Engaging with the critics	Eng5	*Critics and theory*
Awareness of critical context / background	**Eng3**	*Critics and theory*
Confident tone		*Voice*
Deeper / specific analysis - beyond the plot	Arg2; Arg5; Inf2	*Analysis; depth and detail; plot; specific knowledge*
Group 2		
Engaging with the question and quotation	**Eng1;** Eng2; Eng4; Eng5	*Question*
Information - *primary* / secondary / historical / cultural / contextual - balance - relevant to the argument	Inf2, Inf4, Eng3 Inf1	*Evidence; context and period; critics and theory; (quotes)* *Relevance; argument*
Textual analysis (and where it's from)	Arg2; Inf2	*Analysis; (technical – referencing)*
Balance between one-off and close reading of episodes without too much plot		*Analysis; plot*
Creating and sustaining your own argument	Arg4	*Argument; voice; consistency, coherence, focus; progresses, direction*
Group 3		
Engagement with the quotation and question	**Eng1 (**Eng2; Eng4; Eng5)	*Question*
Structure and organisation - coherent, having a direction	**OP1,** Arg1	*Structure; signposting; progresses, direction*
Engagement with the material - enjoyment, confidence - technical analysis? Close awareness? Close reading?	Eng5; Arg5	*Voice* *Analysis; (quotes)*
Originality?	Arg4	*Original*
Range	-	*Range*
Contextual understanding - where the texts come	Eng3	*Context and period*
Paper 1? How literature works, challenging assumptions		-

such as 'engagement' feature in some criteria (in blue), and the wording for a few criteria is closely similar. In Table 11, the Faculty labels are in bold where this is the case. Groups 1 and 3 wrote very similar first criteria: 'engaging with the question and quotation' / 'engagement with the quotation and question'. This is close to Eng1, although the student criteria omit the 'incisiveness' of engagement. This is not surprising, as marks and classifications of essays were not discussed. Crucially, however, the student versions elaborate on the faculty criterion, emphasising the importance of engaging with both the question *and* quotation. As Chapter 6 examines, tackling the quotation is challenging, but through reading example essays, participants came to realise its contribution to writing quality. It is also heartening that Group 3 elaborate on 'engagement with the material' as 'enjoyment, confidence' as well as requiring 'analysis'.

Overall, the workshop groups cover thirteen of the seventeen Faculty criteria. The reflections completed at the end of Workshop 1 elicited students' perspectives on the criteria they had formulated as a group. One item was: 'Did anything surprise you about the criteria that we discussed?**'** While a few students commented on the over-specificity of idiosyncratic criteria, such as that relating to Paper 1, the below are representative of most responses:

> 'Not really as we have mentioned them w/ *[tutor name],* but discussing them in more depth was really useful' (Gemma)
>
> 'Not especially but it did seem that there were ways of achieving engagement with quotes *[i.e. the quotation-question]*, for example, which I hadn't considered.' (Abbie)
>
> 'Not really – they felt like the right ones if that makes sense?' (Rachael)

According to these responses, the criteria formulated from assessing example essays were in alignment with prior learning about writing quality, with tutors, for example. Sixteen of the

twenty-one participants reported having read the Faculty criteria prior to the workshops; although as the quotation from Holly at the beginning of the chapter suggests, they weren't necessarily usable. The above extracts suggest that not only was discussing criteria helpful to students but that assessing example essays raised awareness of other strategies, other 'ways… which I hadn't considered'.

The student criteria are not exhaustive, and four Faculty items cannot be easily identified in the sets. They are as follows: 'intellectual penetration and sophistication of conceptualization' (Arg3); 'accuracy of facts' (Inf3); 'clarity, fluency and elegance of prose' (OP2); and 'correctness of grammar, spelling and punctuation' (OP3). The first two refer more explicitly to subject content and knowledge, which was not a focus of the workshops. They would also not be easily recognised in writing by a disciplinary novice or outsider. The final two may perhaps be taken for granted by students: 'grammar, spelling and punctuation' are usually automatically deployed, and participants were instructed to consider which were particularly important to writing quality.

An important high frequency code appears in the coding and Faculty criteria but not the student criteria: *clarity*. As with grammar and spelling, clarity is perhaps a criterion so taken for granted that it does not warrant mentioning and listing. However, it is a very frequent code, and as Chapter 7 will discuss, became a key metacognitive standard for students, driving adaptations to writing. As the criteria lists were formulated early in the workshops, before students had read peer essays, re-read their own, and received feedback, it is possible that the importance of clarity had not yet become apparent, or least not sufficiently so to note.

While on the whole the student criteria focus more on the procedural aspects of writing than the Faculty set, they still exclude some high frequency codes. These are: *introduction, conclusion* and

quotes. To a certain extent these three elements could apply to any essay writing. Indeed, the tangled web of criteria in the visualisation could apply to any literary essay genre, or even more broadly – it would not be controversial to claim that good writing should always have an argument, be clear, and answer the question, if there is one. However, the particular demands of the exam writing task mean that *how* these criteria are addressed varies from other genres. In interviews, participants perceived that while there were some continuities between exam essays, tutorial and A-level exam writing, the demands of the task required them to adapt. As discussed in the next section and following chapters, *introductions, conclusions* and *quotes* were central to the adaptations made by students. Therefore, while the overall rhetorical goals of essays may be similar, the strategies to enact them are different, and will not appear in criteria lists focused on declarative quality. *Introduction, conclusion* and *quotes* are examples of latent criteria which seem insufficiently significant to include in criteria lists, but are actually crucial to writing development.

In summary, the large overlap between the sets of criteria and the coding indicates that both the intervention and the coding align with the Faculty criteria, suggesting that this mode of peer assessment has good validity, where Faculty values are considered the definition of quality. Moreover, the coding shows that the workshops highlighted the contribution of latent criteria which are potentially important to writing quality, but are too procedural to make into criteria lists. The next section now turns to the adaptations students made to their writing, and the extent to which they drew on workshop learning.

5.5 The challenges of exam writing

Self-regulated learning requires a suitable level of challenge: if a task is too easy, there is no need to self-regulate (Hadwin et al., 2011). Participants in this study found that exam writing posed a raft of challenges, and these are summarised here before moving to how students made adaptations in response.

In Prelims Papers 3 and 4, students choose from 20 or more unseen questions to write three essays in three hours on topics of their choice. In interviews, participants were asked to compare and contrast exam writing with tutorial and A-level essays. In Interview 1, Heather sees the purpose of the period papers as:

> 'To sort of challenge your understanding […] to see how far you do have a good understanding and overview of the literary period 1830 to 1910 – to see how much you understand the different forces and the different texts and the different kind of innovations and traditions and how it links together.'

As Heather goes on to explain, exam essays go further than the 'depth' of tutorial essays, assessing students' overall 'breadth' of understanding, too. Candidates have to take tutorial learning, with each week typically focussing on one author, and synthesise it into topics. A further challenge is that students are required to write three exam essays in three hours. Participants saw parallels in these conditions of production with A-level examinations. However, they considered the demands to be much greater, including far more questions to choose from, having to decide which texts to write on, the lack of certainty around which topics would come up, and writing three essays rather than one or two, covering a wider range of material. Nonetheless, students perceived Prelims writing to be more flexible. At A-level, the need to systematically address Assessment Objectives (AOs) in every paragraph influenced essay structure. As Megan describes, 'you'd like… write sort of

chunked out paragraphs to make sure that you get like all the assessment objectives in the paragraphs' (Interview 2). The extent to which students reappraise A-level ways of writing is discussed in more depth in Chapter 7, on metacognitive standards.

The function of Prelims exams as summative assessments has three particular implications for student writers. First, it shifts the rhetorical goals of writing. While in interviews students reported seeing tutorial writing as an opportunity to explore ideas and learning, as Phoenix puts it, exam writing is a 'result', or a 'consolidation' of that learning. Students need to 'show that you have gained a succinctness and an understanding' (Interview 2), and do not receive feedback beyond marks. Second, the examinations provide an extrinsic motivation to succeed. Prelims are classified as Distinction, Pass and Fail, and do not contribute to the final degree. Nevertheless, colleges often award scholarships to students who gain Distinctions, and students who fail and then fail a re-sit do not progress to the rest of the course. Some participants were very aware of such consequences, and of the élite status of the university. Chloe considers that Prelims test 'all the sort of reasons, they're like, that's why you should have been let in here [laughter] I guess, just sort prove your place slightly, I think [laughs]' (Interview 2). This leads to the third implication: perhaps inevitably, such stakes make Prelims a source of stress. While throughout the data, students express their enthusiasm and enjoyment for the subject matter, the timed conditions and volume of material were considered daunting. In Interview 2, two participants spontaneously voiced that they would not choose to be assessed in such a way, and that the portfolio assessment was more suitable. Amber contrasts the productive 'intellectual pressure' presented by assessment with the intense demands posed by three-hour exams: 'I mean, the pressures of time can do a lot to a person' (Interview 2). As we will see throughout the thesis, such pressures, combined with the goal of demonstrating knowledge, can prompt students to take safer options

which have led to good performance in the past, but may not be as valuable in terms of their current learning.

To prepare for collections and Prelims, students used a range of strategies, the most common of which was memorising quotations, the bread and butter of literary analysis, represented in coding as *quotes*. Without revising quotes, 'I wouldn't have had anything to say', as Ben puts it (Interview 1). In an examination situation, we might expect students to retrieve such knowledge and adapt it to the audience and task as they write: in Bereiter and Scardamalia's (1987) terms, knowledge-transforming. In contrast, the data shows that the Prelims task poses further challenges, and does not reflect a straightforward plan-write-revise process. Instead, the data reflect Galbraith's (2009) dual-process model. All case study participants did some sort of goal-oriented pre-writing planning, ranging from recalling sometimes complex knowledge to jotting down some annotations. Nonetheless, all at some point in the study experienced shifts in understanding and ideas during writing: knowledge-constituting (Galbraith, 2009). This was prompted by the complex question-quotations set by the examiners, which require students to rethink or repackage their prepared knowledge. Indeed, for many participants, their perceptions of the question-quotation shifted, too, as they wrote and discovered new implications. While some participants, such as Heather and Amber, preferred to discover as they wrote, generating new ideas can be disruptive and a change of mind risks undermining the essay's argument or coherence. As Chloe put it, in an exam situation there is not 'the freedom to sort of chop and change' available in more leisurely, exploratory tutorial writing (Interview 2). Partly this is due to time, and partly reluctance to revise – examiners' reports stress the importance of presentation. As discussed next, students tried a range of techniques to manage the complexities of writing, and the workshops opened up new strategies.

5.6 Overview of large-scale adaptations

Prelims provide both incentives and barriers to performing well, and to making adaptations to writing. As the criteria coding shows, students commented on more criteria than are captured in the Faculty or the sets they formulated. Several participants described the workshops as raising their awareness of writing quality, but also commented on the difficulty putting everything into practice:

> 'I feel like when you look at a lot of essays and notice what's good and bad about them you can then see what does it mean to me, how does it apply to me […] Obviously it's hard to change things. But I think at least being aware of it, you know that you can try and work on it, and you're always going to struggle with certain things but I think at least when you're aware of something, you can actively try and develop things, so.' (Tessa, Interview 2)

> 'It's hard, because… I know what I would want to be done differently, but when you're under pressure, it's hard for it to actually happen'. (Felicia, Interview 2)

This section gives an overview of the adaptations students make to their writing during and after the workshops, despite facing the difficulties mentioned above. It also traces the extent to which adaptations can be linked to the intervention, addressing the over-arching research question: (how) does students' self-regulation of writing develop in the context of peer assessment writing workshops?

To address the research question, I conducted analysis on a case-by-case basis. This consisted of three steps: first, identifying self-reported adaptations; second, establishing whether there was textual and other evidence for these adaptations and third, tracing the extent to which the workshops prompted adaptations. For each case study participant, I retrieved data coded at large-scale adaptation in NVivo: both 'new' strategies or changes students reported trying out in an essay, and the 'future' adaptations that students intended to make in other essays. These data

were then cross-referenced with the visualisations and analytical memos to ensure completeness. The self-reported adaptations were collated in tables. For 'future' adaptations, I checked later interviews and other data to see if these had been enacted. Where students reported making adaptations, I corroborated these self-reports with think-aloud protocols, written plans and essays. Next, to establish the extent to which workshop learning had prompted adaptations, I cross-referenced self-reports with the trace data: students' annotations and responses to example essays, as well as feedback data, both given and received. I also ran a criteria coding query to identify where criteria were identified in workshop data and where criteria had been used as standards in the interview and think-aloud case data. To capture textual changes, I compiled a table comparing introductions across the three essays. This process resulted in four summary tables for each case study participant; Megan's are available as an example in Appendix H. I then compiled a master table of reported adaptations for all cases, which was in turn condensed into Table 12 below, which is ordered in rough number of adaptations. As the research question focuses on self-regulation rather than textual development, only intentional adaptations were recorded in the table. This decision is both a strength and a limitation. If learners are self-regulating, then their proactive intentions must be a hallmark of that self-regulation. However, self-report methods such as interviews rely on memory, and instruments eliciting all possible adaptations, however minor. Some more incidental adaptations may not therefore be captured.

Table 12: Overview of large-scale adaptations at Workshop (W) essay, and Interview 2 essay.
Key: (T) – strong textual evidence for adaptation, e.g. word count; (t) – weaker textual evidence for adaptation, requiring some inference; (TAP) – think-aloud evidence; (EE) – example essays; (P) peer essays.

Case	Time	Adaptations (evidence)	Link to workshops
Megan	W; Int 2	Increased focus on question. W2: Referencing keywords. Int 2: deconstructing the quotation; revisiting question throughout essay (both monitoring and through topic sentences) (T)	References to question throughout workshop data. Peer feedback on topic sentences.
	W; Int 2	Experimentation with planning to ensure argument clear. W2: more structure. Int 2: shorter, looser plan at Int 2 (T)	Comments on argument and clarity throughout data (EE, P). *Preoccupation since Interview 1; tutor suggests strategy.*
	Int 2	To ensure she finishes, more focused introduction (T); also limits number of paragraphs (T). Essay shorter but finished (T).	Comments on introductions (EE, P). Runs out of time in workshop essay. Aware of impact of running out on quality (P).
	Int 2	Makes time to re-read and edit essay at the end, adding clarification to introduction (T)	Comments on clarity and introduction throughout data (EE, P). Self-efficacy: stopping before time is up to edit 'strangely a confidence thing'.
Phoenix	W; Int 2	Increased focus on the question, 'using it to my benefit' (Int 2). Int 2: Annotating the quotation (T). More question focus in introductions (TAP; T)	Comments on the question, including annotations made during discussion on the quotation in particular. Reflections on importance of the quotation-question.
	W; Int 2	Increased focus on signposting argument and clarity (t).	Reflections. Comments on signposting and clarity (EE, P).
	Int 2	'Narrowing it down' – being more selective about points in order to finish. Doesn't manage a conclusion, but increased essay length (T) and doesn't think conclusions are so important.	Comments on clarity. Peer feedback on conclusion.
Chloe	W; Int 2	Intentionally limits number of texts and authors. Reduction in number of texts used over study (T). Finishes Int 2 essay, and essay lengths increase (T).	Comments on development (EE). Two peer assessors comment on the need to push points further. Still working on this at Int 2.
	Int 2	Focuses on argument and outlining structure in introduction (t)	Comments on introduction (EE, P); also comments on importance in reflections. At W3, intends to work on intro and topic sentences. Attribution in interview.
	Int 2	More focused on ensuring topic sentences link back (t)	Peer feedback. Attribution in interview.

Case	Time	Adaptations (evidence)	Link to workshops
Nicole	W; Int 2	Referencing – starts to underline book titles (T)	Addition on referencing conventions. Attribution in interview.
	Int 2	More focused on finishing, so moves through points more quickly, changing the structure of her essay so it has more paragraphs (T). Doesn't finish, but essay is longer (T).	Doesn't finish workshop essay. Proposes this solution in W3 reflections; also suggested in peer feedback.
	Int 2	Focus on topic sentences to 'direct argument and make it easy for people to understand' (t).	Comments on signposting throughout workshop data. Peers give mixed feedback, which motivates her to 'be as clear as possible… to avoid any potential ambiguities for markers'.
Felicia	W; Int 2	Tries engaging with the quotation/question more in W essay (t) Partially successful, according to peer feedback and self-assessment. In Int 2, reports 'thinking more about linking stuff back to questions' (t). Again, partially successful.	Comments on question and signposting throughout dataset. Two peers give suggestions. Attribution in interview.
	Int 2	Introduction – takes longer to think it through. Explicit thesis statement. (T)	Comments on introductions (EE). Feedback from two peers suggesting solutions.
Ben	W; Int 2	Greater engagement with the question and quotation. At Int 2, spends time deconstructing question and quotation (TAP). Returns to question once planned (TAP). W and Essay 2 both directly reference quotation (T).	Reflections, comments, additions on importance of question and quotation (EE, P). Also receives peer feedback on his handling. *Tutor suggested deconstruction strategy*
	Int 2	More emphasis on clarity and argument/thesis. During planning, thinks through thesis (TAP), and re-orders material rather than reproducing it.	W essay doesn't have a clear thesis; aware in reflections, confirmed by peer feedback. Reflections, comments, additions on importance of argument and clarity (EE, P). *Tutor suggested planning strategy*
Tessa	W	Planning: Tries to outline a structure (T)	Workshop reflections.
	Int 2	Planning: To set up introduction – annotations; notes argument (T) Introduction evolves, with a more direct thesis statement (T)	Additions made during discussion on introductions. Mixed comments with some praise from reviewers.
Heather	W	Tries planning more and outlining a macrostructure (T)	Planning a concern from Int 1. At W1, argument key criterion; planning 'crucial' to argument (EE).
	Int 2	Engages with question 'a bit better'; intro handles q in more depth (T).	References to question through dataset. Peer feedback on intro.

Table 13: Participants for whom there were no adaptations linked to workshop learning

Case	Time	Adaptations (evidence)	Workshop data related to approach
Holly	Int 2	Same approach (TAP), but structures material differently, dealing with multiple texts in each paragraph. A fourth paragraph (T).	While immediately post-writing she intends two actions related to question, after assessing peer essays, 'I don't think I would try anything different – just more of what I am doing'.
Amber	W; Int 2	At workshop, uses approach as developed for April collections, but managed to write faster. Produces a longer essay at W2 (1328 words). Int 2 essay shorter (1219 words). Intends to write faster next time.	Post-writing, reflects that she wouldn't do anything differently with her approach next time: 'No, I'm happy with it.' After assessing peer essays: 'I would keep the approach consistent & only try to work in critical sources, too.' Decides not to act on the latter.

All students adapted their self-regulation of writing in some way during the study. As Table 12 shows, eight of the ten case study participants reported making adaptations to their writing, which either they attributed to the workshops in interviews, or could be linked back to their comments on example essays and peer writing, their reflections, and their peer feedback. Table 13 summarises the adaptations for the two remaining participants, the bulk of whose adaptations came before the intervention. As discussed below, the workshops served to confirm that their approaches were appropriate and didn't need further change (Section 5.8).

In most cases, participants used workshop learning to make between two and four adaptations to writing. Reflecting the complexity of writing quality, these changes often targeted more than one aspect of writing. The main trend is that many adaptations focused on the *question*, and particularly the quotation within the question, on *argument*, and on *clarity*. These are all high frequency criteria codes, but to achieve them, writers tended to adapt their introductions, or their signposting, procedural criteria not noted in the Faculty or student lists.

Chapter 5 examines adaptations related to task perceptions in particular, so strategies related to *question* are not discussed further here. However, it is worth highlighting how introductions, where students start to show their engagement with the question, changed over the study. A basic measure of change is word counts, as shown in Table 14, with arrow symbols to denote trends over the study. Between pre- and post-workshops, introduction word counts changed considerably for most. Some participants cut opening general contextualising moves, which are more suited to a leisurely tutorial essay (e.g. Amber, Ben and Megan). Others wrote longer introductions, adding moves to refer directly to the question or outline their argument or texts in more detail (e.g. Felicia, Heather, Holly, Nicole, Tessa). The average introduction increased by 15 words, or around one sentence (English students use long sentences; they mirror that aspect of academic practice without issue). The word count of essays also increased across the study, but this is perhaps less indicative of adaptations as students also became more familiar with their material. Essay structure, however, is a more sensitive measure. While most participants started with a three paragraph structure, brought with them from A-level, in two cases, writers intentionally aimed for a different structure in order to ensure they could finish their essays (Megan and Nicole).

Several participants tried to adapt planning process to manage the conflict between knowledge transforming and knowledge constituting (Galbraith, 2009). It is particularly difficult for students to change long-held habits and assumptions that planning entails putting together an outline, which is followed throughout the essay, as Heather's experiences in the next section show. Two participants, Ben and Megan, take up a tutor's suggestion to 'jot down thoughts', as Megan puts it, rather than put together a rigid structure; although in Ben's case, he still needs the three paragraphs he has used since A-level. Chapter 8 examines how Tessa adapts to the challenge of planning as her standards shift.

Table 14: Introduction word counts				
Participant	**Interview 1 Intro**	**Workshop Intro**	**Interview 2 Intro**	**Trend**
Amber	129	76	84	↓
Ben	176	115	107	↓
Chloe	92	94	114	=
Felicia	69	89	130	↑
Heather	76	112	125	↑
Holly	85	135	124	↑
Megan	151	113	109	↓
Nicole	59	171	161	↑
Phoenix	193	216	215	=
Tessa	118	135	153	↑
Mean	114.8	125.6	132.2	↑

Finishing the essay within the time limit was also a common challenge that led to adaptations. For Chloe and Nicole in particular, getting the essay finished was a key goal at Interview 2. Both participants noted in interviews how their essays would be more successful if written as a tutorial essay. But Chloe reflected in Interview 2 that 'you have to be a bit more lenient because you know you've got the time pressure. And so yeah, you have to be more realistic with what you want to achieve'. For both, time was an issue from the beginning of the study – Chloe had run out of time during her first collections in January. The workshops reinforce the impact of not finishing on quality. Nicole notes surprise at the length of others' essays, and Chloe reflects in Interview 2 that:

> 'The amount of essays that we saw that… like sort of weren't quite finished, weren't quite there, and it was just really like sort of 'no!', you knew it could have been so much better if it just had that sort of extra bit'.

As a result of experiencing the frustrations of unfinished essays as a reader as well as writer, at Interview 2, Chloe reported focussing on 'making sure you're getting things done' – pacing herself through the essay. Her essays became progressively longer over the study: from 758 words at

Interview 1 to 890 words for the workshop essay, and 958 words for Interview 2. Similarly, Nicole identified running out of time as an issue in her own work and others. During the workshops, she was surprised by how much others wrote, and her reviewers hinted that her essay seemed unfinished. As a result, in Interview 2, she reported that:

> 'I was more focused on moving forward and trying to… trying to get to the end. There were bits… erm I definitely could have expanded on within the individual paragraphs, but I was like 'no, I've – I've made a point, or at least the majority of the points I wanted to make within this paragraph already, it could be a lot better but this isn't a tutorial essay, I don't have time, and I think it's more productive if I just go onto the next point.'

Nicole first notes this possible strategy after re-reading her essay at the end of Workshop 3 and judging that her paragraphs are lengthy and hard to follow. In the interview, she attributes her enactment of this strategy to a reviewer's advice, which was:

> 'Everything you say is good, so having two less sentences elsewhere would be OK – a conclusion is perhaps going to add more value than two sentences elsewhere in an already good essay!' (Nicole's underlining).

Although Nicole didn't quite finish at Interview 2, there is textual evidence for this adaptation. She wrote her longest essay yet (752 words, compared with 564 words at Interview 1, and 634 words at workshops). In contrast to the usual three paragraph structure, she wrote six paragraphs in addition to the introduction and a conclusion fragment. She considered that with slightly more flexibility in the exam – dividing three hours between three essays – she would have been able to finish. However, the lack of time during term to practice and try out strategies was a significant obstacle to adaptation, as discussed in the next section.

5.7 Barriers to adapting writing

Time was a considerable barrier to adaptation, and this was particularly evident in Heather's case, where there is less evidence of adaptation. It is also worth noting that like others, Felicia in particular, Heather struggled to identify positives in her writing, perhaps influencing her self-reporting of writing development. Immediately after writing, she was still swept up in the ideas, and it was more challenging as an interviewer to get her to articulate writing processes.

To recap, the workshops took place in Weeks 1 to 3 of term, and on Monday of Week 4, a portfolio assessment was released, due Friday of Week 5. The exams were at the end of Week 8. While students had prepared for collections prior to term starting, they reported having only Weeks 6 and 7 to dedicate to the remaining three papers, and colleges ran seminars and tutorials during these weeks. Students scheduled Interview 2 for Week 7, and in one case, Week 8. Heather nearly cancelled her Interview 2 at the beginning of Week 7 because she felt she wasn't yet 'at the writing stage' of revision. However, she used the session to try out her knowledge on a topic. She made some adaptations: she judged that she was more focussed on the question, and analysis of her introduction shows that it is longer and she uses more words to engage with the quotation. Planning was a perennial source of dissatisfaction over the course of the study, and Heather also ascribed issues in example and peer essays to lack of planning. In the workshops, she resolved to take action, but used her usual approach at Interview 2. Heather described her planning as 'look at the question and think of all the different things that I could write', although the think-aloud indicated that she also developed a tentative argument. She explained the source of her difficulties:

> 'I should take a more organised approach, but it's really difficult because even though I feel I should plan, I find it really hard to get my ideas and till I start writing. Like really, really difficult [laughs]'

The ‘should’ here is interesting: planning is an orthodoxy, and certainly one enshrined in the school curriculum, which assumes a ‘chronologically determined’ plan-write-revise process of writing (Myhill and Jones, 2007, p.339). However, for Heather, knowledge constituting leads writing, and she needs to write in order for ideas to flow (Galbraith, 2009). The drawback is that she perceives she loses focus, and ‘a more organised approach’ would prevent this. She intends to do more practice, but also comes up with two alternative ‘solutions’. The first is to choose and plan all three essays before beginning the first essay, as she found text selection decision difficult and was concerned that she had used authors from two of her topics: ‘if it had been an exam paper, I might have struggled more for the other questions’. Heather’s second solution is to ‘leave a few lines in the introduction so afterwards I can go back and really crystallise my argument, ’cause it probably won't be crystallised until the end of the essay.’ A key piece of feedback she intended to enact, according to her Workshop 4 reflections, was ‘making line of argument explicit in intro’, and this solution would enable her to do so without more detailed planning. Like Megan, then, Heather proposes ways to work with knowledge-constituting, rather than trying to impose a heavily organised knowledge-transforming approach.

5.8 What is enough adaptation?

Heather, then, was very much still adapting her approach at Interview 2. However, other students decided that their approach was good enough. Two participants, Holly and Amber, changed their approaches during the study but before the workshops, with peer assessment serving to confirm that their approaches were effective, and thus not in need of radical adaptation.

Holly, who was introduced at the start of the chapter, tells a clear narrative of proactive writing development over the year. While many students found revision stressful, she finds that ‘my

revision's actually been quite fun […] I've been in my room with a cup of tea with some books and some things that I enjoy.' She writes revision cards and sticks them on the walls of her room: 'I call it the goldfish effect, so every time you walk round the goldfish bowl you see something and take it in eventually [laughter]'. Holly describes the workshops as adding to her metacognitive knowledge of her own writing:

> 'I think [I learned] a lot about the way that I write, and the way that I should write […] I realised I needed more evidence, I realised that I needed a tighter structure, I realised how much, exactly how much time I had to play with...'

The workshops, then, add to her metacognitive knowledge, but the longitudinal data show that this knowledge is not dramatically new, and does not in itself prompt adaptations of her writing process. For instance, analysis of her essays show that she uses more evidence – critics and quotations. The first strength she identifies in her Interview 2 essay is that 'it's quite well-informed [laughs]. I think I've got some good primary knowledge, which is something I really struggled with at the start of year, I had [an] argument that wasn't really grounded in the texts.' The workshops, then, underlined the importance of primary evidence, an issue of which she first became aware at the beginning of her transition to university. Rather than offering new metacognitive knowledge about her writing, the workshops reinforce her judgement of her writing, and provide an additional reason to become immersed in the goldfish bowl of material.

Likewise, 'structure' is a long-term preoccupation for Holly, with awareness raised through the reflective exercise she completed after her January collections. For Holly, there are two ways of creating structure in her essays: genre strategies such as topic sentences and thesis statements in the introduction, and the organisation of material. At Interview 1, she reports trying out genre moves such as using two sentences in the introduction beginning with 'this essay will'. During the workshops, she makes seventeen annotations and additions on topic sentences, and then gives

two peers suggestions on using topic sentences and the introduction to create structure. For structure-as-organisation, she makes six annotations and two additions during discussion on the example essays, noting impressions about the ordering of points, and where material might fit best. The macro structure of Holly's essay *does* change between intervention and Interview 2. For the workshops essay, she writes about an author per paragraph, with some topic sentences to highlight connections; she recognises this in one of the example essays. At Interview 2, she writes about her two main authors together in each paragraph, bringing in other texts. The overall pattern of text use is similar to Interview 1, but there is more quoting of primary texts, and she does not use the question-quotation as a structuring device. Such structure-as-organisation in the Interview 2 essay is not achieved through adaptations to the writing process itself. Analysis of the think-aloud and written planning data shows Holly's approach remains consistent. Instead, she works out structure by using practice essays to see 'how the ideas work out', a habit developed at A-level and used in her January collections. She begins with an open book essay, then writes timed essays on generic themes rather than specific question-quotations. This rehearsal of ideas and possible combinations allows her to write quickly, and at Interview 2 she manages to write a fourth paragraph for the first time. Section 7.2 examines this writing in the context of task perceptions, as it seems this method leads Holly to focus on the question less explicitly, thus bucking the wider trend across the study.

Amber, a more reticent interviewee, also had a strong sense of writing development. Her January collections made her 'conscious' of that she needed to adapt her technique: 'just how mindful I had to be of time and how much I was writing and how much just primary material that I had to just memorise.' In Interview 1, she reports making adaptations – moving on from the 'haphazard' structure of her January collections, she tries out a three point structure. While writing the introduction, she discovers her argument and decides to start again, ditching the 'roundabout'

attempt to include a more upfront thesis. She also decides to write on one author only. Amber considers her adaptations to come to fruition a week before the workshops in her April collections essays, which were highly regarded by her tutor. She continues with this system for the workshops and the Interview 2 essay. After the latter essay, she considers writing faster, a strategy that worked out in her April collections:

> 'I think like as far as just getting words down on the page, I took my time a bit. Umm in my last collection, in 0th week of Trinity, I was… really conscious of trying to make my hand go faster than my head. And it – thankfully it turned out OK, but I was really worried that it was just a pile of nonsense as far as the argument goes, but umm here I just really thought about the argument as I was writing it, and so maybe next time I'll just make it go faster now that I know that it kind of in the end worked.

As for Holly, then, the Interview 2 essay serves to see if an argument 'worked', and as with Holly, the argument is refined in response to the question, in Amber's case the ideas in the quotation. Now Amber knows it works out, she can use it again, but with less monitoring – 'mak[ing] my hand go faster than my head'.

What Holly and Amber have in common is that the workshops confirm or validate a self-evaluation of their writing as strong. Holly's evaluation of her essay immediately after writing her essay is that it needs more focus on the question, and more critics. In Workshop 3, she reviews peer essays and re-reads her own. The same critiques still stand, but she notes that 'I am usually quite critical about my own work but I feel surprisingly positive about this essay.' This self-evaluation is in contrast to her usual self-assessments, describes in Interview 1 as 'not a very positive exercise, um, because I… I find it quite difficult to pick up things that I've done well simply because it's me.' Perhaps this 'surprisingly positive' evaluation is because she regards the quality of the peer essays as 'mixed': she finally has the 'benchmark' of quality to make her judgements, and realises they are quite good. Although Holly notes the same limitations as at Workshop 2, she doesn't see the

need to do anything differently next time: 'just more of what I am doing'. To address the need for more focus on the question, rather than adapting her writing, e.g. by adding moves or signposting, she intends to continue 'searching for & answering the right question', reflecting her realisation that another item may have been more suited to her material.

Both Holly and Amber receive positive peer feedback, and annotate comments. Holly annotates each with a 'strong' and a smiley face; similarly, Amber places a tick and a smiley face next to a particularly pleasing comment: 'You demonstrate a very deep and complex knowledge of your chosen authors and texts making your argument and analysis very sophisticated.' Holly considers the feedback to be 'quite insightful'. All three of her peer reviewers suggest adjustments to her introduction – as she summarises - 'work on signposting in introduction'. However, as discussed above, although her Interview 2 introduction is a 'concern', she doesn't follow up the peer suggestions, partly because more overt signposting would be 'still further away from my writing style than I'd like it to be'. While Holly praises the quality of her feedback, Amber finds hers more problematic, as she considers some reviewers may have been too ready to critique for critique's sake. While she felt that overall 'it pretty much matched what I thought', the critical feedback was less helpful: 'each person had kind of a designated paragraph in which they picked out a different flaw that they had clearly kind of hunted down a bit.' She sees these 'flaws' as less important to quality.

Receptiveness to feedback is discussed further in Chapter 8. For both Amber and Holly, the workshops confirmed that their approach was sound, and so there was no need to make further dramatic adaptations. Such positive self-evaluations and feedback may have also served to boost their self-efficacy for writing, which is the focus of the next section.

5.9 Self-efficacy

Self-efficacy arose as an important influence on writing during the pre-intervention interviews in February, and in tutor interviews the following month. This chapter has already mentioned some reasons why students may find Prelims examinations daunting, for example, the time pressures, the range of material, their role in determining progression to the second year. Although Prelims are not as high-stakes as, say, A-levels, the examinations are formal occasions. Over three days, students attend the Examination Schools building to sit three hour exams, and must wear sub-fusc, or full academic dress. Tutors perceived that 'confidence' was important to how students tackled this challenge, and to their performance. Transcription and early analytical memos of case study data also showed how self-efficacy and emotion influenced students' decisions, for example around the questions they chose, as discussed in Chapter 5. According to the Winne and Hadwin (2008) model of SRL, beliefs such as self-efficacy may shape students' self-regulatory decisions, and whether they consider it is worthwhile making large-scale adaptations. Equally, experiences such as those provided by the intervention may prompt students to re-evaluate and update their beliefs: large-scale adaptation can occur for beliefs as well as for metacognitive knowledge and strategy use.

To address this important aspect of writing, a fourth research question was added to the study: 'What is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing?' I designed an instrument based on Bruning et al.'s (2013)'s model of self-efficacy for writing, which proposes three dimensions: ideation, conventions and self-regulation (e.g. planning, revision, time management). Participants completed this survey at the beginning of Workshop 1 (W1) and the end of Workshop 4 (W4).

As the mean differences for most items deviated from the normal distribution, I used the non-parametric Wilcoxon signed-rank test to examine whether there was a difference in median group scores and their distribution between W1 and W4. The results that led to this decision are included in Appendix C. The null hypothesis for each item was there was no difference in rating at Workshop 1 and Workshop 4. The results of the Wilcoxon signed-rank test for each item are presented in Table 15. Alongside the test statistics, effect sizes (*r*) are reported for each item. The statistical power of the data set is limited by the sample size, and so the results are to be interpreted with caution. However, as discussed below, the results for many items reflect the findings of the in-depth qualitative data, which improves their explanatory power.

The largest and most highly statistically significant differences between Workshop 1 and Workshop 4 ratings of self-efficacy were for items related to understanding quality, conventions and expectations around the Prelims exam essay task. The strongest effect size is for Item 6, 'I understand what makes a good Prelims essay', with a highly significant difference between Workshop 1 *(Mdn* = 67.50) and Workshop 4 (*Mdn* = 85.00; *z* = 3.928; *p* = .000; *r* = 0.88). There are similarly strong results for a similar item, no. 9: 'I understand what is expected of me in a Prelims exam essay' (W1 *Mdn* = 67.50; W4 *Mdn* = 80.00; *z* = 3.676; *p* = .000; *r* = 0.82). As a group, then, participants' ratings for understanding the conventions of essay writing improved after the intervention. There was no significant change in medians for the control items on expectations for tutorial writing (Items 13 and 14). This reflects Felicia's perspective that 'reading other people's essays' allowed her to see what such conventions look like in practice:

Table 15: Self-efficacy: Wilcoxon signed-rank test results

		W1 group median	W4 group median	Test statistic	Standard error	Z-Score	Sig.	Effect size (*r)*
1	If the question fits my subject knowledge, I can write a good Prelims essay	75.00	80.00	104.500	17.443	2.551	**.011**	0.57
2	I can write and finish a Prelims exam essay within the hour allowed	70.00	75.00	128.500	19.235	3.145	**.002**	0.70
3	I can use the quotation stated in the question to write my exam essay	67.50	80.00	169.000	22.883	3.649	**.000**	0.82
4	I can plan my exam essay in a way that helps me write the rest of it	55.00	70.00	134.500	21.045	2.756	**.006**	0.62
5	I can effectively manage my time over the three hours of the exam paper	62.50	70.00	109.500	17.367	2.850	**.004**	0.64
6	I understand what makes a good Prelims essay	67.50	85.00	210.000	26.730	3.928	**.000**	0.88
7	I can manage my emotions during exam writing	77.50	80.00	74.000	15.882	1.354	.176	[0.30]
8	In exams, I can plan in a way that helps me generate ideas before I start writing	70.00	70.00	87.000	15.871	2.174	**.030**	0.49
9	I understand what is expected of me in a Prelims exam essay	67.50	80.00	169.000	22.716	3.676	**.000**	0.82
10	I can write an exam essay that addresses the question or prompt	70.00	80.00	142.000	22.702	2.489	**.013**	0.56
11	I can stay focussed for the three hours of the exam	67.50	72.50	123.500	24.645	1.156	.248	[0.26]
12	I can plan my exam essay in a way that helps me organise what I'm going to say	70.00	70.00	128.500	26.681	.881	.378	[0.20]
13	I understand what makes a good tutorial essay	78.50	75.00	62.000	17.454	.115	.909	[0.03]
14	I can write a tutorial essay that meets my tutor's expectations	72.50	75.00	99.000	20.949	1.074	.283	[0.24]
15	After a period of revision, I can find at least three questions on the exam paper that fit my subject knowledge	80.00	80.00	72.000	14.173	1.870	.062	[0.42]
16	I know what to do with the quotation part of exam questions	70.00	85.00	137.500	21.065	2.896	**.004**	0.65
17	I can organise my essay so that it is easy for a reader to follow	65.50	75.00	144.000	22.891	2.556	**.011**	0.57
18	I can write a Prelims exam essay that meets my tutor's expectations	62.50	70.00	142.500	21.030	3.138	**.002**	0.70

> 'It definitely gave me a clearer sense of what err the essays are meant to look like. […] Because having only done tutorial essays, you don't really know how that is then meant to apply, and like how it's meant to be different for an exam essay. So that really helped.' (Interview 2)

There was also a significant, but slightly smaller shift in students' ratings of their self-efficacy to write an essay which meets such expectations. For item 18, 'I can write a Prelims exam essay that meets my tutor's expectations', there is a significant difference and medium effect size from Workshop 1 (*Mdn* = 62.50) to Workshop 4 (*Mdn* = 70.00; $z = 3.138$; $p = .002$; $r = 0.70$). Similarly, for Item 1, 'If the question fits my subject knowledge, I can write a good Prelims essay', there is a modest significant difference and effect size between Workshop 1 (*Mdn* = 75.00) and Workshop 4 (*Mdn* = 80.00; $z = 2.551$; $p = .011$; $r = 0.57$). In Interview 2, Heather explains how greater confidence in understanding expectations made writing the essay seem more 'manageable':

> 'I feel like now I know a lot better what they want from me and a lot better how to achieve it and how, what I need to do in terms of, like, sitting down and planning properly and deciding what I'm going to write about for each one. I think it just made it a lot more manageable. I think it just sort of broke it down a little.' (Interview 2)

Being clearer about what 'they want from me' made it clearer which adaptations she needed to make. As Heather found essay writing more 'manageable' as expectations were clearer, Megan found that the level of writing quality required seemed more 'achievable':

> 'I think... I also learnt a bit of confidence as well [laughs] […] I think like when you're reading other people's essays [pause] I don't know, there was something... about them not being perfect but still being good that makes it seem like achievable, or like you can write an essay of that standard or that you're kind of... because sometimes you think kind of think like, being at Oxford or whatever that everyone's writing these brilliant essays all the time, and so like having the confidence to kind of like say "oh actually, it's like a good essay, but it's still got flaws, too" like makes you think, "well, my essay's also got some flaws, but it's got some good bits as well" so I could kind of write this, if that makes sense.'

Megan here highlights the relationship between self-evaluations and self-efficacy. Reading others' work prompted her to re-appraise her self-evaluations of writing, and as a result, her self-efficacy beliefs, as the 'so' indicates: '"well, my essay's also got some flaws, but it's got some good bits as well" ***so*** I could kind of write this.' If students don't have a 'benchmark', as Holly puts it, against which to judge their own writing, their self-evaluations may be overly negative, and thus impact on their self-efficacy beliefs. This is a particular danger in an environment such as Oxford, where entry is selective, and so students may assume that 'everyone's writing these brilliant essays all the time', and that the benchmark is an undefined brilliance. In interviews, participants tended to be very ready to voice the limitations or problems with the essays, often from the very first question: 'how did it go?' However, it was often hard to get students to find strengths in their work. Felicia reflects on this as follows:

> 'I guess you automatically, as you're writing, even, you're thinking about 'oh, I probably didn't express that quote very good. But you're not thinking, 'oh that bit went good!' [laughs] I don't know, I tend to be thinking more critically as I go.' (Interview 2)

Felicia's default mindset for self-monitoring, then, is critical rather than self-praising. In Interview 2 she still finds it difficult to volunteer strengths of her work, instead focussing on what she 'tried' to and expressing uncertainty about its success. However, the peer feedback she received on the workshop essay – a stronger essay in her view – boosted her self-efficacy: 'it was also quite encouraging that actually I can do this.' Other participants, too, found that feedback highlighted strengths they weren't aware of, and was a source of self-efficacy. This will be discussed further in Chapter 8.

A further area in which there was a significant change in ratings is handling the question, and in particular the quotation included in the questions. For most students, the quotation-question

format was very different to the format of A-level questions. As discussed in Chapter 5, the TAP and Interview 1 data shows that some participants avoided certain questions, not necessarily because of their content, but because of the format. There was a large significant difference and effect size for Item 3, 'I can use the quotation stated in the question to write my exam essay' (W1 *Mdn* = 67.50; W4 *Mdn* = 80.00; z = 3.676; p = .000; r = 0.82). There similarly positive but not as strong results for two further items related to the question: Item 16, 'I know what to do with the quotation part of exam questions' (W1 *Mdn* = 70.00; W4 *Mdn* = 85.00; z = 2.896; p = .004; r = 0.65) and Item 10, 'I can write an exam essay that addresses the question or prompt (W1 *Mdn* = 70.00; W4 *Mdn* = 80.00; z = 2.489; p = .013; r = 0.56). In some senses, it is a little surprising that the results are so positive for this procedural aspect of writing. While *question* is the most frequent criteria code, the workshops did not cover strategies, or use live models dealing with a question, as in other studies on self-efficacy. However, it seems that reading a total of eight examples and peer essays and reflecting on their writing was source enough for a modest boost in self-efficacy for handling the question. As discussed above, it may have also made it seem more 'manageable' or 'achievable', to use Heather and Megan's terms.

Also in the procedural, self-regulatory dimension of writing, the items related to time management – Items 2 and 4 – also saw modest but significant increases between Workshops 1 and 4. As discussed above, students were challenged by time pressures, and Nicole and Chloe both adapted their approach to make finishing in time a goal. For others, the experience of writing a timed essay in Workshop 2 may have provided a boost in self-efficacy, either because they finished in time, or because they saw that others had struggled to finish, too.

The continuing challenges of planning under pressure voiced in the interview data are also reflected in the self-efficacy survey results, which were mixed. For a general item on planning,

there was a highly significant difference and medium effect size (Item 4: 'I can plan my exam essay in a way that helps me write the rest of it'; W1 *Mdn* = 55.00; W4 *Mdn* = 70.00; $z = 2.756$; $p = .006$; $r = 0.62$). This is a rather general item that implies planning is helpful to getting started, to beginning writing, and perhaps the experience of writing the workshop essay confirmed this. However, for two planning items with more specific, precisely-worded purposes, the group medians remained the same. For Item 12, 'I can plan my exam essay in a way that helps me organise what I'm going to say', there was no significant difference between the medians for Workshop 1 (*Mdn* = 70.00) and Workshop 4 (*Mdn* = 70.00; $z = .881$; $p = .378$). For Item 8, 'In exams, I can plan in a way that helps me generate ideas before I start writing', the median remained the same, and the significant difference and medium effect size reflected a shift in the distribution in values between Workshop 1 (*Mdn* = 70.00) and Workshop 4 (*Mdn* = 70.00; $z = 2.174$; $p = .030$; $r = 0.49$).

The results for these items reflect the case study data discussed above around the challenges of planning where a timed task involves both knowledge-transforming from long term memory and spontaneous knowledge-constituting through writing (Galbraith, 2009). There is an assumption that we 'should' plan writing, enshrined from school onwards, and many participants expressed dissatisfaction with their planning. Some participants planned further adaptations to their planning even after Interview 2 (e.g. Heather, Megan). Others, however, found other ways to achieve the purposes of planning, particularly with regard to organising writing even where ideas shifted. This is reflected in the significant difference between Item 17: 'I can organise my essay so that it is easy for a reader to follow' (W1 *Mdn* = 65.50; W4 *Mdn* = 75.00; $z = 2.556$; $p = .011$; $r = 0.57$).

Further items for which the null hypothesis is retained are those on emotional regulation and focus (Items 7 and 11). This is not surprising as the intervention was not designed to address these

areas of self-regulation. In addition, the null hypothesis is retained for Item 5, related to preparation and knowledge. Again, this is expected as the workshops did not cover literary knowledge and preparation techniques.

Overall, then, the results of the survey suggest that participating in the workshops had a positive impact on students' self-efficacy, and in particular for understanding conventions, writing an essay which meets expectations and handling the quotation-question. The mixed results on planning reflect the challenges of the task, and the need to find alternative strategies. There were also mixed results for other procedural aspects of writing, and although a strategies handout was discussed in the final workshop, more follow-up in this area may be desirable, to discuss, for example, alternatives to extensive planning.

Although the emphasis of the thesis is on writing processes, it is important to acknowledge the role of such beliefs in shaping and being re-shaped by writing. Workshop activities allowed students to develop their ideas about quality – Holly's longed-for 'benchmark' – and to appraise and reappraise their own writing in context, at the same time making the task seem more 'manageable', to use Heather's term. The thread of self-efficacy runs through the thesis. Chapter 8 examines the particular impact of feedback on self-efficacy, while Chapter 6 addresses the role of self-efficacy in the task perceptions phase of writing.

5.10 Chapter conclusion

This chapter has given an overview of key themes that underpin the rest of the thesis. The coding shows that students' comments and talk about writing reflected a more complex web of criteria than Faculty or criteria lists. It seems that reviewing example essays made procedural aspects of

writing visible. Such aspects of writing, while seemingly too insignificant to list officially, are important to adaptation, and the next chapter addresses which criteria students used as standards to self-regulate and adapt their writing. Chapter 7 then picks up on the centrality of the question-quotation to writing quality, and examines how students' task perceptions changed over the period of the study. An additional impact of making the web of criteria visible was that exam writing because more 'achievable', more 'manageable' for students. This seems to shed light on increases in self-ratings of self-efficacy, particularly for conventions, and for meeting expectations in writing.

For all participants, assessing and discussing a range of essays raised their metacognitive awareness about writing. Most participants acted on workshop learning to make adaptations to their writing. However, it was important in this chapter to also discuss outlier cases where workshops served to confirm their judgements that they didn't need to adapt their system further. This chapter has introduced the complexities of self-regulating exam essay writing: summative assessment poses both incentives and barriers to adaptation.

Chapter 6
When criteria become standards for writing

6.1 Introduction

As seen in the previous chapter, as self-regulating learners, students choose which aspects of their writing to focus on and adapt. The theory of action posits that taking part in peer assessment workshops enriches students' ideas about quality in writing – their evaluative expertise – and so their own personal standards. In the workshops, group ideas about quality were captured in the sets of possible criteria developed through discussion at the end of Workshops 1 and 2. These sets were in turn synthesised on the strategies handout participants took away together with their feedback. In the literature, it is often assumed that providing students with such criteria translates into learning outcomes that meet these criteria. This chapter sheds light on the learning processes that lead to criteria being adopted – or not – as (meta)cognitive standards for writing. With limited time to prepare for the exams as well as a tight time limit for the exam itself, it would not be feasible, or as Amber highlights, even desirable to try to meet a full set of criteria:

> 'Some [criteria] were clearly a lot more necessary than others. Kind of use of the primary texts, it seems that if that's left out, you're left with like a really seriously lacking essay. Whereas if some of the criteria were left out, it really wouldn't be that big of a deal. Not necessarily kind of something to kill your essay with.'

This chapter explores which criteria students choose to prioritise – the potential essay killers which they can't neglect – as well as those which reflect their own ideas about quality. The chapter addresses Research Question 2: 'Which criteria do students use as standards to self-regulate their essay writing processes?' As Amber describes, not all criteria can become standards for self-regulating writing processes. Taken together, a set of standards makes up a goal (Winne, 2014). In an examination situation, students have less scope to set goals: essays

need to be completed within an hour and students are likely to have performance goals – to produce a 'good' essay. What makes a 'good' essay will vary according to the student's own standards. Therefore, standards are more interesting for investigating individuals' proactive decisions, and there is more likely to be intra-individual variation.

To answer the research question, I first give examples of standards in action, and where standards come from, including writers' awareness of the need to adapt standards from a previous genre of essay: A-level exam writing. Next, I present an overview of coding results on which criteria are translated into the most frequent standards: the essay killers which cannot be 'left out'. Continuing with quantified data, I then compare students' standards at pre- and post-workshops – which standards are brought into emphasis by the experience of appraising others' work. The discussion turns to processes: the use of process-oriented standards which go beyond the criteria observed in others' work, and how standards are used over the phases of learning. I then illustrate key findings with a focus on one case: how Tessa adapts her planning strategies over the course of the study to try to ensure clarity for the reader, a key essay killing standard. The chapter concludes with a discussion of the very utility of criteria.

6.2 Definition and operationalisation of standards

As described in Section 3.8 of the theory of action, the term 'standards' is not used in the sense of achievement standards (e.g. Sadler, 2014). Rather, the personal standards that students use to judge their own writing are self-set and emerge from students' beliefs and experiences of writing. The study uses a working definition of metacognitive standards synthesised from Winne and Hadwin (1998, 2008), Greene and Azevedo (2007) and Poitras and LaJoie (2013). In this study, I define a standard as follows:

> A criterion or belief that is used to judge or evaluate learning, either how it is going in the moment, or how it went (for example, in the self-evaluation part of the interview).

It is very important to distinguish standards from general beliefs about writing. A belief or criterion only becomes a standard when it is used to monitor or judge learning. Writers may hold beliefs that are effectively ideals about writing that they either cannot meet, or choose not to strive towards due to other constraints. A belief is also only considered a standard where it is used in reference to a particular piece of writing or learning, not writing experiences more generally.

Standards come into play from the very beginning of a task. On opening an examination paper and beginning to read, the think-aloud protocols show students making rapid decisions about the suitability of the 20+ questions available. Metacognitive standards often seem to be binary at this point: they determine that a question is suitable or it is not. Participants often reject questions with ‘don’t like that one’ or ‘no’. The most frequent standard code here is *knowledge-question match*: whether the question fits the writer’s prepared subject knowledge. For example, Tessa rejects a question with ‘haven't done drama yet, don't really know enough’ (TAP 1). Another standard may be understanding of the question-quotation: Heather rejects one possibility with ‘don’t really understand that quote very well’ (TAP 2). Writers may also have strong preferences for the type of question: for example, at TAP 1, both Phoenix and Chloe both reject quotation questions: ‘don’t like discussing quotations’ and ‘kind of instantly drawn not to do quotation questions.’ (Chloe, TAP 1). It could be inferred that they need to feel comfortable with the format of question; the standard could also be emotional.

Once the task is chosen, some participants have very explicit standards, using checklists, either written or held in mind. Here, Holly describes how with time to spare, she checked the Interview 2 essay against her standards:

> Holly: And when I go back through, I just wrote down the names of the critics that I've got in there, so I've got [counting] 1-2-3-4-5-6-7, and then these are my primary quotations, so I've got 10 in there [laughs]
>
> NU: Wow! That's quite packed [laughs].
>
> Holly: Yeah, and that's in theory, that's how much I should be able to remember for three essays going into the paper. So I mean I was quite pleased about that.

Here, the standards seem to be 'remember and get in *n* critics' and 'remember and get in *n* primary quotations'. While Holly doesn't give the explicit number, we can infer that at 7 critics and 10 primary quotations, she has met her standards. These standards reflect the first criterion she names in her response to the interview question on what makes a good Prelims essay: 'it should have a high level of primary and secondary information'. Holly shares that primary quotations are 'something I've struggled with, I can do the secondary quite easily, but I think the primary, I tend to read the text, get some themes from it, and then forget […] I could just use a couple of words here just to crystallise that example.' Counting her quotes and critics is a strategy to ensure that she meets this standard, then, and emerges from metacognitive knowledge of her reading and use of primary texts.

Holly is the only participant recorded counting up elements of her essay that meet her standards, and other writers tend to have a more fuzzy sense of what is 'enough', particularly for elements of writing that are not countable. Megan's description of what she learned in the workshops provides a good data-driven definition of how writers might experience standards:

> Umm [pause] I think I learnt [pause] err like how, I don't know, it's not quite *how* you write a good essay, but like… where to like kind almost like, not start with it, but kind of like **what kind of things could be ticking over in the back of your mind umm as you're writing** and I think I kind of learnt how to respond to the questions… better because I'd never like kind of been faced with a question before coming here [i.e. Oxford] that had like a quotation and a thing, and then kind of like looking at other example essays helps you to understand like how you can bring like two facets together. (Megan, Interview 2, original emphasis in italics; my emphasis in bold).

Standards, then, are 'things [that] could be ticking over in the back of your mind… as you're writing' that can add up to a good essay – an overall goal. Reading examples of work gives Megan ideas and possibilities for standards she could adopt, and helps her develop task perceptions for handling a type of question she hadn't experienced pre-Oxford. As this chapter will discuss, the workshops gave students an opportunity to develop their standards. Before turning to discussion of which criteria became standards, and how standards changed over the course of the study, the next section deals with other sources of learning about writing that can influence the standards participants hold about writing.

6.3 Where do standards come from?

Other than the workshops, three main sources of standards can be traced in the data: tutorial essay writing, tutor feedback, and school writing (A-level). However, these standards have to be renegotiated or adjusted to be relevant and feasible for exam essay writing.

In the Oxford tutorial system, students write at least one tutorial essay each week for discussion with their tutor and one or two peers. Although this system offers Oxford students more writing experience than those at other universities, writers in the study were very aware of differences between the tutorial essay and exam essay genres, and that tutorial essay standards were not immediately transferable. Participants saw tutorials as an exercise in learning, whereas exam essays 'show that… you have learned a lot in the year and that you can use that knowledge in kind of unique and interesting ways when presented with quotations you haven't seen before' (Amber, Interview 2). The exam essay requires different uses of knowledge, as Heather explains:

> 'I think a tutorial essay is possibly a space where you get to explore a very niche idea and you can go into lots of detail on it and you have quite a few days to write it and you do lots of reading specifically for it. Whereas an exam essay slightly tests what you already know and you slightly have to show everything

> you know about a specific area. […] I think a tutorial essay can just be about depth. In an exam, you have to have range and depth.' (Heather, Interview 2)

The exam essay is more demanding, then, with the need to go beyond 'exploring' ideas to 'show[ing] everything you know' requiring students to use both 'range and depth'. Writing under 'intellectual pressure' as Amber puts it in Interview 2, participants see a need to adjust other standards. Here, Tessa recognises this in view of the memory constraints posed by the task:

> 'In the exam you obviously have to reduce yourself to some basic critics that you can remember, otherwise you're never going to remember everything' (Tessa, Interview 1)

Writing under timed conditions also necessitates a different process to the more leisurely tutorial essay. Some writers fit descriptions in the literature of revisers rather than planners – those whose dominant approach to managing writing is write first, revise heavily later (Torrance et al., 1999; Galbraith and Torrance, 2006). Here, Nicole recognises the need to adopt a new standard for Prelims writing:

> 'Yeah I think [pause] er **having kind of clarity of what, what I want to say before I start writing** was um… cos it's something you don't have to think about so much when you're writing a tutorial essay cos you'll write something and then you'll write it again and again and again. And you can't really, you don't really have time to do that in the collections essay, which is a problem. Um [pause.] So yeah I kind of felt that, like no **you need to really have an argument before you start** and um don't kind of just wander off and then that way if you kind of like have a concept of what you're doing in your head then you won't be going back and changing and adding asterisks and stuff like I did throughout my collections essay.' (Nicole, Interview 1; my emphasis)

Rather than working out her argument by 'writ[ing] it again and again and again', the constraints of the timed essay means that Nicole has to learn to 'have an argument before you start'. She assumes that adopting this standard should mean that she can avoid 'changing and adding asterisks and stuff'. It is interesting that revision is not desirable behaviour for writers;

this perhaps reflects time constraints but also the emphasis in school writing on presentation, which is also emphasised repeatedly in Oxford examiners' reports (e.g. 2016).

While college teaching was not a focus of this study, tutor feedback and advice is an important influence on standards. At Interview 1, many students reported how they tried to act on advice and also feedback from college formative exams – collections – sat around a month earlier:

> 'I was more conscious of the importance of argument, because that had come after my last collection.' (Chloe, Interview 1)
>
> 'Um I think I planned for longer than I did in my collections because um yeah, basically [laughs] because my tutor said that I should spend more time planning [laughs] so I tried to do that. (Megan, Interview 1)
>
> 'So with my tutor at the moment we've been working on like drawing out implications at the end of the argument - and I didn't [laughs]' (Felicia, Interview 1)

These examples show how at Interview 1, these writers try to readjust their standards, focussing on argument, and allowing more time to plan. Felicia's example shows that in this instance, 'drawing out implications' was a desirable but not yet achievable standard, due to time. This is perhaps more appropriate to tutorial writing, the genre for which she was learning this move. Nearer exams, revision classes are mentioned, and the interview data allow the interactions between workshop and college learning to be traced in individual cases.

While participants perceived university writing to be different to school writing, their experiences of A-level essay writing were still influential, as Baker (2017) also found in her ethnographic research across subjects. With limited time for exam preparation – three weeks once coursework was submitted – participants continued to use revision strategies developed at school such as memorising quotations and developing mindmaps. A-level criteria, too, continued to influence standards, although unlike Baker's (2017) participants, in a renegotiated way. Here, two terms into her university career, Tessa is able to reel off the A-level assessment objectives:

> 'I think the key thing that I had at A-Level is you had the different assessment objectives. So [laughing] they're just drilled into you, so it always gets you to think about AO1, sort of the way that you write, your language and your prose style and then AO2 was always umm attention to detail, so analysing specific words. AO3 was critics and AO4 was context.' (Tessa, Interview 1)

Students describe how at A-level, they were encouraged to show evidence of each assessment objective, covering each objective in every paragraph they wrote: 'you'd like... write like sort of chunked out paragraphs to make sure that you get like all the like assessment objectives in the paragraphs' (Megan, Interview 2). Thus, it seems that A-level assessment criteria strongly influenced students' standards for writing, and this influence remains during university. Pre-workshops, Felicia discusses how her A-level learning is helpful for exam writing, but needs to be reappraised. Note how like Megan, she, too, uses the phrase 'drilled into you':

> 'So you get drilled into you such a structure at A-level and then coming to uni it's like, oh no you have to get rid of all of that and start again and start a different way. But now actually as I go along, I'm like actually no, the stuff they did say in A-level was kind of relevant, I just have to do it in a different way. So A-level was super structured so we had like all your different AOs, AO1, AO2, AO3, AO4. And as you went through, at each point you had to address each AO. […] I think stuff that I learnt at A-level that I didn't really realise I'd learnt has suddenly like come back, so like analysis. And oh actually, no, I do know how to analyse and like we had prepared that stuff even though when I first started term, I was like 'I can't remember how to do it!' [laughter]. So it [A-level] does, did prepare to some, an extent analysing text but also bringing in context as well. (Felicia, Interview 1)

Felicia, then, can see continuity between A-level and exam writing, particularly in the key skills of analysis and context, but this has to be used differently, rather than 'super structured' to evidence assessment objectives. Felicia's retelling of AOs resonates strongly with Torrance's (2007) description of 'criteria compliance' in Further Education – where giving evidence of criteria becomes the key objective of learning. Training in criteria compliance continues as an influence at Interview 2 for Felicia, who feels that 'I know it's only a week to go – but umm I haven't quite got a like a good structure in place yet'. For her, A-level still provides a paragraph structure and standards for writing in timed conditions:

> ‘Whereas I didn't really do that for tutorial essays, for exam essays I do think more about umm like having a quote that I've learnt, and then linking it to a critic, linking it to critical theory or linking it to context and then linking it back to the wider umm topic, theme, argument. So I think that structure kind of carries on, for me anyway.’

Other participants, too, find that A-level continues to be influential. Ben describes how he continues with the same structure: ‘I still have three paragraphs. I just – it just has to have three paragraphs [laughs] because it's just the structure and I don't think I can deviate from that, it would stress me out too much [laughter]’. However, he no longer calls on A-level criteria during writing:

> ‘When you're doing an A level that's very much [inaudible] the criteria thing, you've consciously got these criteria that you're fulfilling, but the step up, at least at the moment, or at least… it's such a step up from A-level means that they've kind of all been integrated into the way I think anyway. So there's no need to think, 'oh I'm not writing enough critical... points in this paragraph' because there will inevitably be a critical point in it because of the way that my argument functions.’ (Ben, Interview 2, my emphasis)

For Ben, A-level criteria no longer have to be ‘consciously fulfill(ed)’ and so no longer represent key standards. A-level provides a foundation, and the criteria have been subsumed into ‘the way I think’ as he has developed his writing. This hints at how some standards such as ‘critical points’ become unconscious – they become an unquestioned part of the writing process. Standards can also be based in routines and processes. At Interview 1, Phoenix began his essay by analysing a quotation – not the quotation in the essay title, but a ‘simple’ quotation with which he was already familiar. This was a strategy used in his tutorial essays, but also allowed him to get into a familiar routine:

> ‘It allows you, particularly allows me to get into the… the rhythm of writing, because if you start with simple stuff like language, style and form, it’s stuff that we’ve been doing since like aged 12, so that gets you into the rhythm straightaway. You’ve got a point down that you can explain, and then from there it moves on quickly.’

Starting with this other quotation is an emotion regulation strategy, allowing Phoenix to get started on a daunting task (Webster and Hadwin, 2016). The underlying standard here is to feel comfortable, to 'get into… the rhythm of writing'. By doing this, Phoenix can also meet criteria-based standards which he has held since 'aged 12': 'language, style and form'. This phrasing is close to the A-level assessment objective, and his description in the final sentence also reflects the point, evidence, explanation structure which is widely taught for response to literature as young as Year 7 (Ofsted, 2009), or Year 6 (Usher, 2013). Meeting a 'simple' standard of explaining a familiar quotation, is motivational – it allows him to tackle the more difficult aspects of writing. It also allows him to do something he enjoys – analysing 'juicy quotations'. It is easy to see why writers find it difficult to move on from such familiar routines and standards that allow them to get comfortable and use an enjoyable and fluent skill. To an extent, writers will never be able to 'get rid of' A-level and 'start again', as Felicia put it, but standards can be made visible and renegotiated, or perhaps even become inactive.

6.4 Which criteria are used as standards?

Section 5.2 introduced the criteria codes identified in workshop data, interviews and think-aloud protocols (TAPs), and this section will introduce which of these students used as standards for self-regulating writing.

As might be expected given the discussion above, some criteria codes are continuations from participants' A-level experiences, if only in name. Codes such as *analysis, critics and theory, context* and *quotes* reflect Tessa's descriptions above of A-level AOs, although *critics* is expanded to *critics and theory* as participants' talk also encompasses the literary theory introduced at university. Although there is a good overlap between the criteria coding and the Faculty criteria (Section 5.3), only one code was named the Faculty nomenclature – *original*. At Interview 1, several participants voiced concern about producing an *original* essay, as Phoenix

says: ‘I know myself and a few other people get quite worried sometimes about trying to get that [originality of thought] into an exam’. However, by Interview 2, this is renegotiated into something more feasible and useful as a standard for writing:

> ‘I think it [a good exam essay] has to have an – an element of originality of thought, umm but not as, not necessarily as much as people like feel they have to because I mean... I've never understood why or how people may expect Prelims students to come up with a brilliant new thesis in your first year, when OK, you know the mass critical corpus out there, it's an established fact that everyone's known for centuries but for us it, we're coming to it for the first time, and it's very much new for us, so it has to have an element of the original, but not afraid to just grapple and put what's already extant in your own way, or in your own view. I think the originality comes from the way you view it and express it, rather than whether it's groundbreaking.’ (Phoenix, Interview 2)

This example shows ‘originality of thought’ is renegotiated into ‘the way you view it and express it’, and made more manageable and useable as a standard. Other participants reinterpret it as ‘ambitious’ (Nicole, Interview 2), and as providing novelty: ‘you can come up with quite a lot of the same points for lots of the same authors, umm so maybe… Interesting sort of combinations or interesting ideas’ (Chloe, Interview 2). Phoenix’s view on originality reflects how self-efficacy is embedded into how participants reinterpret and operationalise official criteria – Phoenix considers that students have to be ‘not afraid to just grapple’ with knowledge, and this can be said for intimidating criteria as well.

The purpose of Table 16, on the next page, is to compare the criteria identified in workshop data and talk about writing in pre- and post-writing sessions with those actually used by writers as standards. The first column gives frequencies of references in example essays data, the second combines references in interview and TAP data on both occasions. The final column shows the frequency of references evidencing use of criteria as standards in TAPs and interviews on both occasions. Green and blue highlighting is used to show the top ten most frequent criteria for each column; grey highlighting denotes the top fifteen most frequent

Table 16: References to criteria codes

	References to criteria – example essays data	References to criteria – interview and TAP data	References in interview and TAP data where criteria used as standards
C1 Analysis	103	43	28
C2 Argument	239	117	78
C3 Choice of authors or texts	70	2	1
C4 Clarity	194	49	38
C5 Comparison and linking	90	17	12
C6 Conclusion	139	42	38
C7 Consistency; coherence	58	27	25
C8 Context	130	51	28
C9 Critics and theory	298	107	61
C10 Depth	21	16	8
C11 Development	118	24	20
C12 Evidence	127	6	4
C13 Finishing	2	13	11
C14 General; broad	28	3	2
C15 Introduction	230	40	36
C16 Knowledge…	71	77	35
C17 Nuance; complicating	24	23	17
C18 Original	4	21	5
C19 Plot	30	2	2
C20 Progresses; direction	23	11	9
C21 Question	542	214	125
C22 Quotes	125	70	43
C23 Range; scope	54	41	28
C24 Relevance	67	18	15
C25 Signposting…	164	32	23
C26 Specific, selective	2	28	19
C27 Structure	69	47	36
C28 Style	116	7	3
C29 Style – concise, precise	17	15	6
C30 Technical - referencing	48	8	5
C31 Technical accuracy	137	9	7
C32 Terminology	8	11	8
C33 Voice, positioning…	116	8	0
C34 Word count; length	20	16	12
Total number of references	3484	1215	788
Mean	102.47	35.74	23.18
Standard deviation	101.30	38.94	22.28
Median	70.5	22	16

criteria. Inferential statistics have not been used as the number of each references in each column are markedly different, which would make it difficult to make meaningful comparisons between datasets. Also, the example essay data is essentially evaluations on finished essay products, while the TAP and interview data contains evaluations on processes as well as products. For think-aloud, only the data for the five participants who opted to participate at both Time 1 and Time 2 is reported here.

The highlighting shows that there was considerable overlap between the most frequent criteria used to judge example essays and the most frequent criteria used to talk about and used as standards to regulate writing. Six of the top ten criteria codes are consistent across the example essays, interviews, TAPs and are used as standards: *argument, clarity, conclusion, context, critics and theory*, and *question*. *Analysis*, *introduction*, *quotes* and *signposting, topic sentences and paragraphing…* are all within the top ten or fifteen for the three categories of data.

However, some codes are used more frequently to judge example essays than to talk about one's own writing and as standards. Two codes appear in the top ten example essay criteria, but not in the most frequent fifteen for the interview and TAP data. The first, *evidence*, is a broad code which seems to be operationalised as *quotes*, a highly frequent standard which is also in the top fifteen most frequent example essay codes. The second is *technical accuracy*, used for errors, e.g. in tenses, spelling. This code is frequently used in readers' annotations and includes corrections made on essays as they read. The overall low frequency as a standard suggests that writers may take *technical accuracy* for granted due to their competence, or lack time to check their work. *Read over*, a related process standard discussed later in the chapter, appears as a standard writers would meet if they could, but is relatively infrequently enacted due to time constraints. It is also

worth noting that as discussed in the next section, 6 of the 7 uses of *technical accuracy* as a standard occur at Time 2, suggesting that writers come to see it as more important over the study.

Interestingly, some criteria are more often used as standards and in interviews and TAPs than to comment on example essays. These are: *analysis, consistency and coherence*, *nuance; complicating, structure,* and *range and scope*. This is perhaps because readers may make a one-off overall judgement on these criteria, but writers may refer to them as standards multiple times during writing.

Knowledge, understanding and ideas is also more frequently used as a standard than an example essay criterion. This is perhaps explained by the emphasis in the workshops on commenting on how the essays were written, not their content. A separate code include in the 'other' scheme, *content,* with 187 references, is used for references where readers are commenting on the content of the argument or of the ideas discussed, for example disagreeing with an interpretation of a text. In contrast, *knowledge* is concerned with how the essay is written and the impression the reader develops of the quality of the knowledge.

It is interesting to note which codes are rarely used as standards. Some criteria are perhaps too broad to be helpful when writing, conveying readerly impressions rather than writerly strategies. While *style* and *voice, positioning of the author and opinion* are in the top fifteen most frequent codes for example essays, *style* is only used as a standard on three occasions and *voice* not at all. These codes are often used for the reader's impressions, and their broad nature reflects the emphasis on holistic judgement in the workshops. *Voice* is related to argument – strong voice or positioning can be the overall result of the writer paying attention to argument as they write. *Choice of authors or texts,* too, with 3 standards references, is also a broad criterion that seems to

be an initial impression on the novelty or interest of the writer's combination of texts, rather than a standard that needs is required throughout writing.

Two of the least frequent standards reflect writers' idiosyncratic beliefs and preferences about quality in writing. Discussing *plot* is often highlighted as a weakness in the example essays, and only one participant – Amber – includes it as a strength of her own writing (on novelists). *General; broad* is also not often a desirable attribute – *specific, selective* is far more frequently used as a standard by writers. Phoenix, however, sees the broad focus of his Interview 2 essay as a strength.

Thus, while a core set of criteria is used to judge others' and one's own writing, some criteria which are useful for others' writing are not necessarily helpful for regulating one's own writing. This seems to be partly due to individual preference, as we will see later in the chapter. It also reflects that some criteria are perhaps too broad to translate into standards, or relate to a reader's overall impression of an essay. The next section will consider whether and how criteria and other standards shift in use over the course of the study.

6.5 Do standards change over the course of the study?

Table 17 shows references to standards at TAP/Interview 1 and TAP/Interview 2, ordered by differences between the two times. The table includes the 14 codes capturing process standards that emerged inductively from the TAP and Interview data. These were created where existing criteria codes did not adequately capture the standard being used. The code *idiosyncratic* is used for standards which only occur in individual cases. Again, only the think-aloud data for the five participants who opted to participate at both Time 1 and Time 2 is reported here.

The final column reports the differences in frequencies of standards between TAP/Interview 1 and TAP/Interview 2. I ran a Wilcoxon signed-rank test in R to verify the statistical significance of the overall differences (Field et al., 2013). The null hypothesis was that the differences in frequencies between TAP/Interview 1 and TAP/Interview 2 were symmetrical. This was rejected ($W = 238$, $p = 0.034$), indicating that the differences are not symmetrical and that some codes changed in use as standards more than others. The largest differences in frequencies were all positive. These were for *question, clarity* and *signposting,* followed by *critics and theory, keywords* and *task perceptions as standards.* These are likely to drive the asymmetrical differences between Times 1 and 2, but it is not possible to run further inferential tests to confirm this as there are not sufficient data points.

Question had the largest increase in frequency between TAP/Interview 1 and TAP/Interview 2, with 35 references. Participants also used *task perceptions as standards* and *keywords* more frequently at Time 2. These frequencies are in line with the findings reported in Chapter 7, which report how students became more focussed on the question and using their task perceptions throughout writing. *Complicating the question* was not used a standard at Time 1, but there were 4 references at Time 2. *Understanding the question* dropped in frequency between the occasions. Together with the findings of Chapter 7, these observations suggest that participants' task perceptions were moving beyond simply trying to understand the question. Using their *task perceptions as standards* and drawing on *keywords* are concrete strategies for showing focus on the question.

Table 17: Standards used at Time 1 and Time 2

	Time 1 (T1)			Time 2 (Time 2)				
	TAP 1	Int 1.	Σ	TAP 2	Int. 2	Σ	Σ T1, T2	Diff. T1, T2
C21 Question	9	36	45	3	77	80	125	35
C4 Clarity	0	8	8	1	29	30	38	22
C25 Signposting	0	2	2	0	21	21	23	19
C9 Critics and theory	5	21	26	1	34	35	61	9
SRL8.6 Keywords	0	1	1	0	10	10	11	9
SRL8.11 Task perc. standards	0	13	13	1	21	22	35	9
C15 Introduction	0	14	14	0	22	22	36	8
C34 Word count; length	0	2	2	0	10	10	12	8
C32 Terminology	0	1	1	3	4	7	8	6
C16 Knowledge…	1	14	15	0	20	20	35	5
C18 Original	0	0	0	0	5	5	5	5
C30 Technical - referencing	0	0	0	0	5	5	5	5
C31 Technical accuracy	0	1	1	0	6	6	7	5
SRL8.2 Complicating Q	0	0	0	0	4	4	4	4
SRL8.3 Ideas	0	2	2	0	6	6	8	4
C22 Quotes	5	15	20	2	21	23	43	3
SRL8.1 Better planning	0	13	13	0	16	16	29	3
SRL8.7 Knowing going to say	1	3	4	1	6	7	11	3
C8 Context	2	11	13	1	14	15	28	2
C12 Evidence	0	1	1	0	3	3	4	2
C3 Choice of authors or texts	0	0	0	0	1	1	1	1
C20 Progresses; direction	0	4	4	0	5	5	9	1
C24 Relevance	2	5	7	0	8	8	15	1
C26 Specific, selective	0	9	9	0	10	10	19	1
C2 Argument	3	36	39	1	38	39	78	0
C11 Development	0	10	10	0	10	10	20	0
C14 General; broad	0	1	1	0	1	1	2	0
C19 Plot	0	1	1	0	1	1	2	0
C29 Style – concise, precise	0	3	3	0	3	3	6	0
C33 Voice, positioning	0	0	0	0	0	0	0	0
SRL8.5 Interesting	3	12	15	0	15	15	30	0
SRL8.8 Knowledge-Q match	2	14	16	1	15	16	32	0
SRL8.10 Read over	0	3	3	0	3	3	6	0
C7 Consistency; coherence	0	13	13	0	12	12	25	-1
SRL8.4 Idiosyncratic	1	11	12	3	8	11	23	-1
SRL8.9 Q characteristics	0	6	6	0	5	5	11	-1
C6 Conclusion	1	19	20	0	18	18	38	-2
C27 Structure	2	17	19	1	16	17	36	-2

SRL8.12 Time	0	13	13	0	11	11	24	-2
SRL8.13 Understanding of Q	0	4	4	0	2	2	6	-2
C13 Finishing	0	7	7	0	4	4	11	-3
C28 Style	0	3	3	0	0	0	3	-3
SRL8.14 Use the plan	0	5	5	0	2	2	7	-3
C1 Analysis	0	16	16	0	12	12	28	-4
C17 Nuance; complicating	3	8	11	0	6	6	17	-5
C5 Comparison and linking	4	5	9	0	3	3	12	-6
C10 Depth	0	7	7	0	1	1	8	-6
C23 Range; scope	2	15	17	3	8	11	28	-6
Σ	46	405	451	22	552	574	1025	

Other differences between standards use at Time 1 and Time 2 suggest increased attention to one's reader. The largest percentage increases were for *Signposting* (950%) and *Clarity* (900%). In the coding guide, *clarity* includes the writer's impressions of the clarity of their own essay, and their intentions to make their essay clear and easy to read. *Signposting, topic sentences and paragraphing* is used for specific references to the strategies the writer uses to guide the reader through the argument. Taken together, these results suggest that at Time 2, participants were more reader-oriented in their choice and use of standards, both in terms of ensuring an overall impression and the strategies used. There is also an increase in the use of *Introduction* as a standard, albeit more modest (57.1%). Outlining the argument in the introduction is a common strategy. Interestingly, *argument* itself remains static in frequency between the two occasions, suggesting that while having an argument is already accepted as an important essay element, making one's argument clear to the reader is a particularly salient criterion for this exam essay task. In other words, it seems students are developing both knowledge of how an argument should be set up in the exam essay genre, and the associated strategies. Finally, as noted above, for *technical accuracy* there is positive increase between Time 1 and Time 2 – one of the highest percentage differences at 500%. This, too, suggests concern for one's reader.

There are also modest increases in *quotes* and *knowledge* as standards between Times 1 and 2 (15% and 33% respectively). This could imply that participants are more focussed on showing the reader or examiner their knowledge. However, it may also reflect the timing of the meeting – participants used Time 2 as exam preparation, scheduling the task around a week or even just a few days before the Prelims examinations. As discussed in Chapter 5, learning quotations was a common revision strategy reported at Time 2, and few participants had prepared for Task 1 as it was taken in term time.

Decreases in some codes between Times 1 and 2 suggest that students are more confident of the task demands of the exam essay genre. In interviews, *depth* was cited a key difference between tutorial and examination essays. The drop in *Depth* (C11) suggests that participants have a better understanding of how much depth is feasible in an exam essay. At Interview 1, adding more points or material was a common response to the question asking what writers would do with 10 minutes more. In contrast, by Interview 2, there is an appreciation that the demands of this task afford only selective depth. *Range*, too, drops, which suggests that students are more confident in how many texts or authors they need to cover; *comparison and linking* is also less of a concern. *Time* and *finishing* drop slightly, suggesting that participants are more used to writing an essay per hour. *Analysis* and *nuance* may also become taken for granted, and are standards that don't need to be mentioned so frequently. In contrast, *technical – referencing* increases from 0 to 5 references between Times 1 and 2. Fieldnotes show that in the workshops, referencing was the only issue consistently not raised by students in discussion but instead brought up from the teaching notes. The increase in *referencing* as a standard implies that managing referencing practices under time and memory constraints is an implicit aspect of the task that became better understood over the course of the study.

6.6 Summary of coding findings

Analysis was also conducted to identify which standards were used in particular phases of writing. This is included in Appendix N as the frequencies are relatively low and most standards were used as self-evaluations of writing during the interview. Two key messages emerge from the patterns of use. Firstly, *question* was used as a standard throughout writing, with increases in some phases at Interview 2; there was also an increase in use of *task perceptions as standards*. These findings will be discussed further in Chapter 7 together with the other task perceptions findings. Returning to the main focus of this chapter, *clarity* and *signposting* were also used throughout the phases of writing, and increased in frequency post-workshops. These codes are both used during enactment and together with *question* and *structure*, are the most frequent codes for what participants reported as *new today.* The use of these codes are standards during writing supports the observation made above that making one's argument clear to the reader is a particularly salient standard for this exam essay task. Again, *clarity* seems to be the overall impression, while *signposting* provides a concrete strategy for making the argument explicit.

In summary, the results in these tables shows that a core set of criteria codes were frequently used in the workshops, to talk about students' own writing and as metacognitive standards for self-regulating their writing. There seems to be consensus across the group that these are essay killers, if neglected. The results also hint at how some criteria are more useful for regulating learning than others, and some are prioritised. A key change in standards across the study was that participants used more reader-oriented standards. Such standards were related to setting up the argument and making it clear, indicating that students' standards became more adapted to the exam essay genre. The discussion now returns to the case study data to illustrate how standards shifted for one writer, Tessa.

6.7 Focal case: Tessa

The quantified summaries of coding showed how *clarity* became more frequently invoked as a standard over the course of the study. Tessa has been selected as the focal case for this chapter as *clarity* plays an important role in motivating adaptations to her planning. As the visualisations in Section 5.2 showed, other criteria codes contribute to *clarity*. Tessa's case is representative of the group in that she works towards *clarity* by adjusting and focussing on other standards.

Across the group, *question* was also a key driver for adaptation, and this is discussed in more detail in the next chapter, on task perceptions. Unusually for the group, Tessa had followed an A-level specification which included quotation-questions, and so had been taught at school to approach the quotation as well as question. Thus, this chapter focuses on her planning, which was a challenge for many participants across the group. Tessa has also been selected for discussion as her subject knowledge was more consistent across the two interviews, and so less of a confounding factor than other participants. At Interview 2, the topics Tessa had prepared at this point didn't come up on the paper, so she chose a question for which her knowledge was 'vague' but 'I want to see where I'm gonna go with that, what do I know so far'. Her goal here is similar to Interview 1, where she writes on theory she had been discussing in tutorial that morning.

Table 18 shows how references to *clarity* increase in Tessa's data over the course of the study, in line with other participants. In Workshop 1, clarity of argument was a key stand-out criterion in determining what was crucial to quality: 'The second essay on India was the strongest as it had the clearest argument and most impressive critical engagement.' Tessa prizes engagement with criticism, and again in Workshop 3 after reading and assessing peer essays, she reflects:

> 'I was impressed by the range of critics and quotes remembered but was often a bit confused by the arguments presented. The question was not always specifically addressed so it was difficult to follow.'

Reading peer essays underlines for Tessa that memorising critics and quotes is important, but not sufficient to quality. As a reader, she experiences confusion, and many annotations and comments to peers relate to *clarity*.

Table 18: References to *clarity* in Tessa's data

Data source	References to *clarity*
TAP 1	0
Interview 1	2
Responses to example essays	2
Annotations and additions	10
Reflections – Workshops 1 and 2	1
Reflections – Workshops 3 and 4	5
Feedback written to peers	9
Feedback received	9
Interview 2	7

6.8 Tessa's planning at Interview 1: 'it's just sort of…. a mess of random ideas'

At Interview 1, Tessa expresses dissatisfaction with her planning, terming it 'just sort of… a mess of random ideas'. Her description of planning as recalling texts, 'key theory that I thought I could relate into it', and some 'key scenes' and then 'just sort of a basic… plan of where I was going to go' matches the think-aloud protocol transcript. However, the interview doesn't mention how much she spent grappling with the question. Analysis of the TAP transcript, excerpted in Appendix O, shows that Tessa returned to the question repeatedly. Of the 330 words of the transcript from the point where Tessa opted to write on the question until where she started writing, 109 were used to re-read the question or repeat keywords, and 50 to interpret or grapple with the question. Tessa's written plan, too, included in Appendix P, contains phrases from the question, and notes where she seems to be making sense of parts of the question.

It seems, then, that task perceptions dominated pre-writing – Tessa interprets the question and generates ideas in response, but doesn't seem to organise them. The introduction of her essay (Appendix Q) also shows her grappling with the question and linking it to theory. In the interview, she describes her planning as 'getting the ideas flowing that I needed at the beginning', but considers that 'I think [laughing] I need to get better at planning because I tend to just write random ideas and not in any sort of order except as I randomly go through, go through, put arrows'. Contrary to this self-report, she does order texts in her written plan, and the essay itself follows this order. However, this 'order' of texts is not enough:

> 'I get [laughs] halfway through it and then think "Wait, where am I?" [laughter] I've got all these points but I don't know where to go with it or you have another idea, umm, that you wrote at the beginning but you don't quite know how to link it to the idea you've just said. So you don't want to just insert a random, I don't know because you want your essay to be cohesive, and have an argument running through it. Err, and I think it's… the difficulty of managing to find a more general argument so you might have several ideas but it… I think you want ...like you don't want to just sort of ramble on with all sorts of different ideas. You want a nice clear [laughs] argument where the ideas all link in to each other or relate to each other. And I think that can sometimes be quite difficult to do in the pressure of an exam because you've only just seen something for the first time and so… You obviously give yourself a bit of planning time but you think [laughing] 'oh no, I've got to start writing now' [laughs] and then yeah.'

In Tessa's view, planning should, then, enable her to 'find a more general argument' that she can use as a standard throughout writing, so that when she pauses to self-monitor – '"Wait, where am I?"' – she knows where to go next. This would help her meet a standard which is perhaps more of an ideal at present – 'you want your essay to be cohesive… you want a nice clear [laughs] argument where the ideas all link in to each other'. It seems that generating the 'general argument' before she starts writing isn't yet feasible due to the 'pressure' of the timed situation. She allows herself some planning time, but standards related to time – "'oh no, I've got to start writing now"' – pull more strongly. Tessa also here articulates a common problem seen

throughout the data – that in addressing a complex task seen for the 'first time', students experience the dual processes of writing – of 'all sorts of different ideas' being generated as they write, but the timed task assuming a linear, ordered process (Galbraith, 2009). In this particular essay, the challenge of balancing knowledge transforming and knowledge constituting is intensified because Tessa chooses to write on new material. She considers that in her prepared collections essay, she had a clearer idea of 'where I wanted to go, implications of arguments because I'd thought about them a lot more over the Christmas holiday'. Despite this, her dissatisfaction with her planning also seems to extend to her collections rather than just Interview 1.

6.9 Workshops: Planning for structure

During the workshops, Tessa tries a different approach. The self-report data shows how Tessa made connections between the example essays and her own writing. In Workshop 1 reflections, Tessa makes inferences about writers' processes and highlights that something from example essays which she does in her own writing is 'lack of planning and straying into a random mix of knowledge I have on an author', citing a previous essay: 'realism in my collections'. For her, the most important criteria that the group generated related to dating critics, and to 'link to the specific words in the quotation' and '[an] argument that is consistent throughout', echoing her comments on the need for a 'general argument' in Interview 1. When she writes her own essay during Workshop 2, Tessa reports that 'I felt like I planned better and was more aware of what the successful structure is in a prelims essay'. A new strategy that she tried was:

> 'planning how the essay would progress rather than just writing down random ideas that came into my head. This gave me good structure, but I felt a bit constrained with critics who didn't fit my specific argument'

She also reports that:

> 'I stuck to my original plan / bullet points but this is unusual. Normally a discussion of one author will lead me on a tangent related to them (even if it's not relevant).'

Overall, Tessa considers that 'I felt like I planned better and was more aware of what the successful structure is in a prelims essay'. Interestingly, although she covers most points in her plan, they are do not follow the numbered order. Taken together, this data suggests that Tessa felt the pull between knowledge constituting – being 'led… on tangents' – and knowledge transforming – following a linear plan and only including the 'relevant'. She reports the essay's strengths as 'its structure. I think it is well-planned and clear'. So, it can be inferred that for Tessa, 'well-planned' or better planned equals a structured essay, during which she sticks to the plan. Re-reading her essay after writing feedback to peers in Workshop 3, Tessa considers that 'I think my essay has a clear structure […] I think I planned well'. In Workshop 4, her peers also seem to agree that she has good structure, but offer other suggestions for improving clarity. For example, Tessa notes that:

> 'Reviewers 1-3 point out that my second paragraph could be made more explicit by adding an extra sentence. This would make my argument clearer and help to further develop its implications, but I didn't realise this needed to be explicitly stated when writing'.

Structured planning, then, is one strategy that can help Tessa achieve clarity, but the workshops also open up other options, such as using signposting at the end of paragraphs. Tessa's approach to planning in Workshop 2 can be characterised as 'structured'. However, post-writing, she notices notes a drawback of this: feeling 'a bit constrained'. She addresses this in her writing for Interview 2.

6.10 Tessa's planning at Interview 2: Structure as restrictive

At Interview 2, Tessa's description of her pre-writing approach initially sounds similar to 'getting the ideas flowing' at Interview 1. She reports writing down authors and points, and identifying 'key bits of the quotation that I thought I had to engage with and then just thoughts I was having immediately'. She then plans some content for her introduction and wrote down 'stuff that I often forget, so critics, context, sort of general literary theory'. Thus, planning acts as an aide-memoire to ensure she includes meet key standards: critics, context and theory. However, in contrast to the workshops essay, Tessa does not then take a 'structured' approach to planning:

> Researcher: So you had a paragraph structure before you started?
>
> Tessa: Umm I more had sort of just general annotations around the question and then just before my answer, wrote some brief points for the introduction in particular, because I think I tend to like to arrange my ideas in my head and on the page, for what I want to put in the introduction because then if I make a really clear, explicit introduction, it sets up my argument so then from that I can see what I'm thinking of generally, following what I wrote before.

The focus of her planning, then, is to set up her introduction so that it is 'clear, explicit' and helps her to monitor as she writes. The ideal of generating a 'general argument' from Interview 1 appears to have become a realistic standard for the planning phase of writing. Although Tessa did not complete a think-aloud protocol at Interview 2, her written plan provides traces of this process. She seems to sketch out her argument in relation to the question:

> Argument → calm, cheerfulness, disinterested objectivity
> x C19th mimesis → 20thC interiority
> with Freud
> ↙ relates to this idea of realism and
> impressionism from Eliot
> 'it is before all to make you see" (Conrad)

As shown in Appendices P and Q, most of her plan is used in the essay. The first points stated in the plan can be seen in the introduction of her essay, although she saves the point about Eliot and the quotation from Conrad until the second main body paragraph. Her starting point – a selection from the quotation – shifts from the 'calm, cheerfulness, the disinterested objective' identified in her plan to 'the dialogue of the mind with itself'. In the interview, she identifies the 'calm and cheerfulness' as influencing new ideas in the essay, and she quotes this part of the question in her third paragraph to introduce the discussion of war poets.

Tessa sees the introduction as a key strength of her essay: 'it sets up the argument quite well' and contributes to her essay being 'quite clear and quite focused'. However, a clear introduction isn't in itself a sufficient strategy for ensuring clarity. To manage the challenge of ideas shifting as she writes – Galbraith's (2009) knowledge constituting – Tessa states that with ten more minutes she would have reviewed her introduction and conclusion to check them for consistency and framed points where she 'strays' as an 'implication, not completely something new'. Indeed, in her essay conclusion, Tessa echoes her introduction, but doesn't explicitly refer to the question, returning to a critic – Whitworth – rather than the Arnold quotation cited in the question. Perhaps at this point, she is focussed more on knowledge transforming rather than the new knowledge she is constituting in response to the quotation. However, in the interview, she is aware of how this 'straying' may come across as 'confusing' to the reader.

Despite making adaptations to her planning at Interview 2, Tessa is still not entirely happy. In response to the interview question on what she would do differently next time, she proposes a further strategy:

> 'I think… just maybe plan a little bit more at the beginning with quotations, because I think sort of, certain ones were coming to me as I was writing, and I used them but I don't know if they were necessarily the best quotations.'

Tessa considers that if she includes quotations in her plan, she can then decide whether to use them, 'whereas I think when you're in the middle of writing you go with it and because your argument is then dependent on that, you're kind of stuck with that quotation now'. Including quotations in her plan offers flexibility rather than a fixed order of material to cover. She characterises planning as:

> 'Planning is just me getting my ideas down and sort of understanding where do I want to go, what I think about this and how I can write it. But not necessarily exactly what I'm going to write because I almost feel that you get a more original reading as you're genuinely… because as you write it, it's almost like your thought process and what you're genuinely thinking about, and whereas I think if you plan a lot in advance, a) it takes up a lot of time that I don't think, I personally couldn't do that anyway, but also… I don't know, it kind of seems a bit sort of restrictive, like I just worry that if I planned it too much in advance, it would just be me inserting another essay or something that I already know, which wouldn't necessarily **engage** with the question, it would just be 'oh this is what I already know, let's shove that in'. So.' (my emphasis)

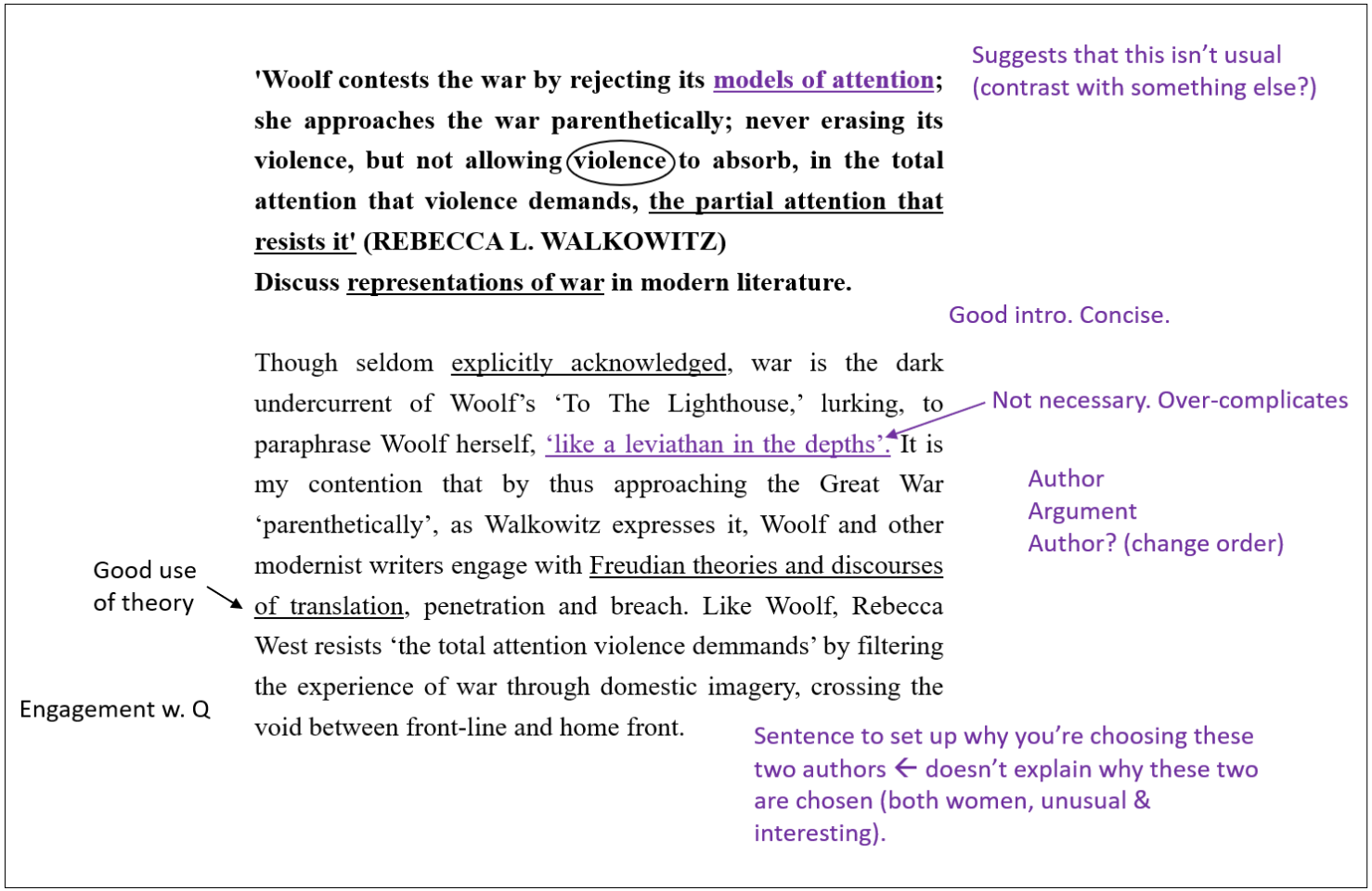
The standard guiding Tessa's planning here seems to be 'understanding where do I want to go' – her general argument, her possibilities for quotations. This seems to offer her some protection from the pitfalls of knowledge constituting – for example, being 'stuck' with a particular quotation rather than selecting the 'best' which in turn would offer affordances for analysis or argument. However, her standard is not as definite or concrete as 'exactly what I'm going to write'. Aiming for 'how I can write it' opens up the possibilities of knowledge constituting to meet the standard of 'get[ting] a more original reading' by writing about 'what you're genuinely thinking about'. Time still seems to be a pressure and a reason to discount planning 'a lot'. However, it seems that offering a 'genuine' response to the question is a more important standard – allowing for knowledge constituting rather than the knowledge transforming or even knowledge telling that relies on 'this is what I already know, let's shove that in' (Bereiter and Scardamalia, 1987).

Tessa's standards then, shift from planning for structure to planning for the introduction. While Tessa identifies her introduction as a strength of her essay, *structure* itself is not identified as a standard at all in Interview 2. The only reference to the term 'structure' in the sense of essay structure relates to the discussion on A-level, and how while the essay macrostructure is similar, the introduction uses different genre moves – 'I think with my A-level essays I'd just go straight in… I wouldn't necessarily outline my argument in the introduction'. Traces of Tessa's growing awareness of possible alternative moves for introductions can be found in her workshop data. In Workshop 1, she notes that:

> 'I was surprised that whilst the introduction should make the argument clear it does not need to be very detailed. Must remember you are under timed conditions'.

Tessa here notes that standards for the introduction can be different to those for other essays, and are adjusted 'under timed conditions'. Her copious annotations on example essay introductions, added to during discussions, show increasing awareness of possible moves. Figure 13 gives Tessa's annotations (in black, on the left) and additions made during group discussion (in purple, on the right) on the introduction to Example Essay 4, which participants assessed during Workshop 2, before writing their own timed essay. Working on her own, Tessa underlines and circles key parts of the question, and notes the essay's 'engagement w[ith] Q[uestion]'. She also notices the writer's 'good use of theory'. The additions made in purple show how the group discussion raised her awareness of quality: it is a 'good intro. Concise'. She labels and challenges the way the writer sandwiches the argument between two authors, and highlights a possibly missing move – that the writer could have explained why this choice of authors is interesting.

Figure 13: Tessa's annotations and additions on Example Essay 4, Workshop 2



With regard to her own workshop writing, though, Tessa barely mentions her introduction in post-writing reflections, all four of her reviewers commented on it, two with unequivocal praise for its clarity and nuance. This shows how self-report data may not pick up all of students' emerging standards. Changes in moves for introduction may quickly seem normal and either not come to mind, or seem worthy of comment. However, textual analysis, presented in Appendix Q, shows how Tessa's introductions evolved over the course of the study. In all three essays, she connects with the question early on in the introduction, but the ordering of her moves varies. At Interview 1 and Workshops, she begins with a point about critics or theory and then brings in the question; at Interview 2, she approaches the question a little more directly by analysing what part of it 'signifies', then connecting this to the literary context and Freud. In Workshops and Interview 2, the final sentence of her introduction states her intentions for the essay in the first person: 'I wish to explore how…' and 'I wish to reexamine how'. This move is first seen in two of three of her April collections essays, with the phrase 'I wish to demonstrate' used in both. Outlining the argument directly using 'I' is not new to her, but the workshops perhaps cement the importance of this move, and also the ordering of moves, starting with the question first.

6.11 Other standards for writing: Tessa's 'boxes'

Tessa's approach, then, can be characterised as shifting towards planning for the introduction rather than planning for structure. In her Interview 2 appraisal of the essay's strengths, as well as highlighting the introduction and its contribution to the essay's clarity, she identifies other areas of strength in her essay as follows:

> Umm I think the introduction is quite, it sets up the argument quite well. So I think… like the strength is that it's quite clear and quite focused and it's written quite clearly, and just sort of throws in everything that ticks the box, like ah, critic,

> contexts blah blah blah, have I got an argument? Is it clear? So I think it's just quite explicit and clear in that way, so the reader wouldn't get too lost [laughing] hopefully.

Tessa again emphasises clarity and the reader's experience, but seems almost blasé about her other standards in this section of the interview – her 'boxes'. Like Holly's counting up of critics and quotes, Tessa's 'everything that ticks the box' seems quite routine. The boxes are unpacked further in a follow-up question:

> Researcher: So what are the sort of boxes that you're thinking of?
>
> Tessa: Umm basically just do I engage with the quotation, have I set out my argument, have I explained which authors I'm using and why, have I got critics, have I got context or theory, so. Critics, context and theory, you don't necessarily need to include all three, I don't think but I think it's – am I showing that I've got this basic critical awareness and that I'm not just… Like showing basically some sort of secondary reading and incorporating it.

At first glance, Tessa's Prelims boxes or standards seem quite straightforward – the self-questions are closed and seem to imply the presence or absence of each item, and there is a continuation from A-level. However, Tessa's 'boxes' are used differently at degree. Her use of context, critics and theory is less prescriptive – whereas she and other participants describe A-level writing as requiring a targeting of each assessment objective in every paragraph, here she considers that 'you don't necessarily need to include all three'. Elsewhere in the interview, she emphasises the importance of 'mak[ing] sure that it's relevant and that it backs up your point' rather than at A-level, sometimes including context because 'you just had to tick the context box'.

Tessa's 'boxes' are placeholders, then, for more complex ideas. Note that she uses the term '*engage* with the question' – in the extract discussed above, over-planning is seen to inhibit engagement with the question as it could lead to 'just… me inserting another essay or something I already know'. In her discussion of Prelims versus A-level expectations, Tessa describes

engagement as the distinction between the two exams. She considers A-level as testing ''do [students] know what this is'', whereas Prelims goes beyond:

> 'I feel like with umm a Prelims essay, it's more like 'do they know it', obviously yes, that's good they know this [laughs], and it's like why, why do they think this, so it's a lot more questioning, it's not just – it's almost as if have they actually engaged with this, have they done extra research on their own, and do they really know the texts, it's not just a superficial understanding. It's – I feel like the keyword is engagement, because it's knowing... umm something kind of a little bit more, and what you think about it.'

For Tessa, 'engagement', then, means showing a deeper knowledge that also encompasses the writer's own position and research. Interestingly, 'engagement' is a key word used in the English Faculty criteria. Tessa doesn't use the term at all at Interview 1, but there are 22 uses of the term or its near forms at Interview 2. As with the example of Phoenix's interpreting of 'originality of thought' in Section 7.4 above, it seems that Tessa has developed ownership of the term, and translated it into something that can be enacted, if not easily articulated. Tessa's 'engagement' seems to go beyond boxes, and blend several of the criteria codes together – it seems to involve voice and knowledge and inferring much more beyond the written essay.

6.12 Is it enough to tick 'boxes'?

When in the interview Tessa summarises what she learned from the peer assessment workshops, the ideas of box ticking and engagement arise again:

> 'Umm I think the most important thing that I found is sort of just a general impression of an essay. Because I think it's so difficult, there's no such thing as saying, and I feel like you convince yourself that there's one way to write an essay, and that I must include x, y, z and then I'll get a First or whatever [laughs]. And I feel like it was really important for me to see a lot of different essays, and see the different ways that people did things […] And I quite like, what was it, holistic criteria that we did? I quite like that because, again, it was like what stands out to you that

> makes it good, rather than simply does it tick this box, because I think when you're writing to just a tick a box, it's – it's not necessarily the same, it doesn't necessarily show engagement.'

The 'general impression' that Tessa develops through reading a range of essays helps her to challenge the idea, then, of 'one way' to write an essay, and enables her to be perhaps more flexible with her approach and standards. It is interesting how she sees 'writing to just tick a box' as not sufficient in itself to essay quality; 'engagement' seems to come with the impression on the reader – something that stands out beyond the ticked boxes. Her description resonates with Sadler's idea of tacit knowledge – that quality can only be experienced, not taught (1987; 1989; 2010). She describes the workshop experience as enabling her to understand the criteria in the context of her own writing: the workshops enable her to develop declarative knowledge about her own writing.

> 'Just generally when I've been writing just looking at the criteria and like the criteria that we devised that makes a lot more sense to us rather than because I think sometimes well, basically the criteria that we are given from the examiners we don't really understand [laughing] what they, very vague and I feel like when you kind of look at a lot of essays and notice what's good and bad about them you can then see what does it mean to me, how does it apply to me, the sort of things that you notice you do wrong yourself. Not that it's wrong, but not necessarily that the best parts of your essay and in another's essay you can see how they do it and you kind of become more aware of your own.'

Reading essays makes not only criteria such as 'engagement' more relatable, but enables her to develop metacognitive awareness, and so to self-assess her own work more fully. This developing awareness chimes with other participants – for example, Phoenix talks about broadening his awareness beyond solely content and ideas: 'now I'm actually focussing not just on what I write, but how I write'.

Other participants, too, talk about going beyond 'boxes'. Nicole characterises the workshop learning she has tried to put into practice as:

> Just the things we talked about before, the criteria and… not the boring stuff, exactly, but the necessary things that… kind of get swept aside when you're just thinking about the – the big thinking and all the interpretation that you do. Yeah. Clear and bringing across that you know things, yeah.

As with Tessa's boxes, Nicole's standards are readily condensed into shorthand: 'the criteria'. It is notable, though, that again ideas of clarity come up and that she singles out the need to be 'bring across that you know things' to the reader. However, the 'criteria' seem almost in opposition to 'big thinking' and 'interpretation', or at least secondary to them – they are in themselves 'necessary' but are easily 'swept aside'. This tension between criteria and 'big thinking' is also voiced earlier in the interview when I ask Nicole what makes a good Prelims essay:

> 'I don't know… is it better to be, to stick to like a criteria [sic] or to try and… I think you're supposed to be able to do both, but we never really talked much about having really interesting, different ideas, but **maybe if you can just have your own, not be actually seeking that, but just… be original and write what's interesting, I think that would come across**, if you find it interesting in your essay, while sticking to all the kind of criteria we have of writing style, and signalling what you're doing, timekeeping and all the rest of it.' (Nicole, Interview 2; my emphasis)

For Nicole, then, it is perhaps most important to 'just… be original and write what's interesting' and allow this to 'come across' to the reader. It's interesting how she described this as unintentional, or rather not being original for the sake of trying to show originality – 'not be actually seeking that' during writing. This again seems resonant of the knowledge constituting process in writing – like Tessa, Nicole describes preferring to revise rather than plan heavily, and she develops new ideas as she writes. However, Nicole acknowledges that she would still need to at the same time 'stick to' criteria, again highlighting two which would shape the reader's experience: 'signalling what you're doing' and 'timekeeping' to ensure the essay is completed.

Focussing on overt standards is, then, a check on creativity, as with Tessa's experiment with a more structured approach to planning.

The tension that Nicole voices between creativity and criteria implies that peer assessment may always be to an extent 'convergent' (Torrance and Pryor, 1998; Pryor and Crossouard, 2008) or even 'conformative' (Torrance, 2012). While I tried to set up the peer assessment so that students were empowered to develop their own ideas about quality, the writing situation is ultimately a rehearsal for a relatively high-stakes assessment situation, in turn involving a power imbalance. Nicole's mention of 'timekeeping', which she and many others struggled with during the study, serves as a reminder that the writing situation is highly stressful and not how participants would necessarily choose to be assessed. 'Timekeeping' is essential to task compliance (Sadler, 2010), and if neglected is as such an essay killer, as Amber would put it. However, it is also important to reader experience. Taking part in the workshops enabled participants to experience for themselves the impact – and often frustration – of reading an unfinished essay. The timed nature of the task imposes an extra, challenging constraint, but ultimately pure knowledge constituting and 'big thinking' is only possible if writing exclusively for oneself. As Virginia Woolf wrote of her own process while writing *The Waves*, 'I am casting about all the time for some rope to throw to the reader'. The 'ropes' used to guide readers in exam writing are perhaps more conformative than in modernist literature, but they are necessary all the same.

6.13 Chapter conclusion

The coding results in this chapter gave an overview of how writers adapted their standards during the study, becoming more focussed on the question, and on the reader's experience, and in particular clarity. Tessa's initial ideal was for her essay 'to be cohesive… you want a nice clear

[laughs] argument where the ideas all link in to each other'. The data show how 'a nice clear argument' became an enactable standard as she adapted her writing, choosing to experiment with her planning. The peer assessment workshops also opened up other options, such as alternative moves to use in introductions, and using signposting at the end of paragraphs. Tessa's chosen changes to her writing process also address her standard of offering a 'genuine' response to the question, allowing for knowledge constituting during writing rather than relying on purely knowledge-transforming, or even 'this is what I already know, let's shove that in'. At Interview 2, she is not entirely satisfied with her approach and has ideas to adapt it further.

The coding results also showed a core set of criteria codes were useful to participants both for judging example essays and for self-regulating participants' own writing. Although no criteria were pre-determined, and students had the latitude to develop their own ideas about quality, participants' own standards do converge with the examination criteria, at least in the terms they use. As with Tessa's use of 'engagement', participants can reappropriate terms and make them serve learning. Nicole's description of 'big thinking' and originality shows how student writing has the potential to be 'transformative' as Torrance (2012) puts it, pushing against ideas and creating new knowledge. However, within the constraints of the assessment situation, there will always be the need to 'stick to' certain criteria which will ensure task compliance: the intended reader is an examiner. Timed writing does constrain creativity.

This chapter has also begun to discuss how aside from the core 'essay killers', students' standards are personal, and to an extent idiosyncratic. This will become increasingly apparent in the task perceptions chapter, next, and in feedback chapter, where standards shape how students interpret and use the feedback they receive.

The findings in this chapter also raise two issues which will be examined further in the discussion chapter. The first issue is theoretical: in the shift to more reader-oriented standards, participants were trying to anticipate audience needs. This connects with the writing theory: from knowledge-transforming to anticipating not only how ideas are represented, but a reader's interpretation of these ideas: knowledge crafting (Kellogg, 2008). In her writing, Tessa balances knowledge-transforming and knowledge-constituting modes of writing (Galbaith, 2009) with a focus on the reader. However, it is not clear whether the writing task affords sufficient possibilities for a move towards knowledge crafting, particularly because time constraints limit opportunities for revisions, such as the edits Tessa would make to her introduction to ensure consistency.

The second issue is more practical. Nicole's wish to 'just… be original and write what's interesting' highlights a potential limitation of the workshops: the exclusive emphasis on writing rather than the interaction between subject knowledge and writing. This was to ensure a consistent focus for research purposes and because a core of common subject knowledge could not be assumed as participants came from mixed college groups and had studied very different sets of authors. As a result of the emphasis on composition, we didn't discuss whether or how interesting and original ideas could outweigh other considerations and make an essay truly outstanding. As Nicole implies, at a high level of expertise, writing and knowledge cannot be easily separated, as knowledge is produced through writing. This issue will be dealt with in more detail in the discussion chapter.

Chapter 7
Task perceptions

7.1 Introduction

In Chapter 5, *question* was shown to be central to the complex web of criteria that makes up writing quality. Chapter 6 showed how writers adapted their standards to be more question-oriented. Engaging with the question is no simple matter, however, and this chapter examines the crucial task perceptions phase of writing, in which students 'themselves create the problem they solve' (Flower and Hayes, 1980, p.23). In the theory of action, developing appropriate task perceptions is essential to both the writing process and the quality of the written product (Section 3.7). Writers bring to essay questions their experiences and beliefs, and also, importantly, their literary knowledge and ideas. With few explicit instructions, the questions need to be interpreted on implicit and socio-cultural levels (Hadwin and Winne, 2012). It is in this interaction that task perceptions are formed, and the rhetorical problem defined. If students' task perceptions are mismatched with their readers', they are unlikely to produce a text which will be judged as successful. In other words, they are unlikely to achieve task compliance (Sadler, 2010).

Task compliance – 'the congruence between the type of response stipulated in the task specifications and the type of response actually submitted by a student' – is a perennial issue in higher education (Sadler, 2010, p.543). In the Oxford context, examiners' reports are available for advice on the VLE, although data from the end of workshops questionnaires indicates that under half of the participants had accessed reports (9 of 21 participants). The reports highlight the ever-recurring issue of task compliance, as in this extract on the Finals Shakespeare paper. The meaning of the coloured text will become apparent shortly.

> As in previous years, the weakest work struggled for relevance, often focusing on a single word in the quotation or ignoring the quotation altogether and **'downloading' a tutorial essay instead of candidates considering how best they might adapt their knowledge to the demands of the question.** Weaker essays were also descriptive and there was a disappointing number of character studies that ignored all sense of context, even at the simplest level, such as where the plays under discussion fitted into Shakespeare's writing career. **All instances of irrelevance were heavily penalized, including some essays that had otherwise excellent qualities.**
>
> (ELL, 2014, p.27; my emphasis)

Ideally, then, during the task perceptions phase of writing, examiners would like students to 'consider how best they might adapt their knowledge to the demands of the question'. At Finals, the question *is* the quotation – students are not also given a question, as at Prelims. To examiners, students should not perceive the question as reducible to 'a single word' or even to be 'ignor[ed]… altogether', as this leads to them reproducing a 'downloaded' essay. The phrase 'downloaded essay' is contemporary, but the complaint of task incompliance is long-standing, as this Examiners' report on the 1935 Finals shows. The coloured text highlights concepts parallel to the 2014 report.

> 'Many candidates handicapped themselves seriously by entering on an examination with, to all appearances, **a large stock of ready-made answers. Instead of trying to think out the new question on the spot, they often seemed to rely on their recollection of old essays.** Few candidates seemed to be aware of **the heavy penalty which irrelevance invariably carries with it.** Direct approach to a question with the support of suitable comment, however meagre, must always secure credit, but evasion of what is asked may throw doubt not only on a candidate's knowledge but also on his qualities of mind, **and so may detract from the credit accruing for his better work.'**
>
> (OUA E/R/1/5, p.26, cited in Currie, 1994, p.129; my emphasis)

Aside from the more inclusive stance and language used by the 2014 examiners' report, the similarities between the two are striking.[1] The papers have changed in scope since 1935, but the demands of the examination are similar, and an eighty-year discourse of 'relevance' persists.

[1] I would like to note that the 2014 extract was not specifically chosen for its similarity. I selected the 2014 report as it would have been the most recent available to the 2015 cohort of participants. I then ran a search for 'relevance' in the report, and selected this passage as it explains the examiners' stance in detail.

Bringing a large amount of literary reading to the task, students continue to face challenges thinking out the new question on the spot', and 'adapt[ing] their knowledge'. Interpreting the quotation could be one issue stymying candidates' attempts to adapt to the question. Prelims candidates are offered a scaffold: quotations are accompanied by a question. However, the presence of this separate question could in itself complicate the task. From a transfer perspective, making the task seem more like other, prior writing tasks, like tutorial or even A-level essays, may prompt over-zealous transfer, or the use of old routines which appear to be helpful to achieving goals, but actually prevent students noticing and engaging with the new (Schwartz et al., 2012). In Prelims, candidates often focus on the familiar – the question – instead of the new – the quotation, as discussed in the 2015 report on Paper 3, Literature in English 1830-1910:

> 'The best answers used the quotation accompanying the question to significant effect, taking from it key terms or ideas which then sustained the entire argument. It is disappointing, however, that a considerable number of candidates did not address the quotations on the question paper and it is hoped that this can be strongly encouraged henceforward: often the boldest and most interesting work is to be found in a candidate's engaged and questioning reaction to those unseen provocations. Weaker answers often ignored the specificities of the question and offered what read like prepared answers.'
>
> (ELL, 2015, p.11; my emphasis)

Engaging with the quotation, then, can enhance the quality of the response. However, for a first-year student new to the task, the Faculty's stance on whether using the quotation is necessary for task compliance is not clear-cut. Responding to the quotation is 'strongly encouraged', but failure to do so does not explicitly carry a penalty, although in the final sentence, irrelevance is again cited, but this is related to 'ignoring the specificities of the question'. It is clear, nonetheless, that using the quotation can lead to writing that goes beyond mere task compliance and is 'interesting', embodying disciplinary values in its 'engaged and questioning reaction'. However, without the guild knowledge and evaluative expertise (Sadler, 1989) that examiners and teachers accrue

through assessing a range of work, students may not understand how the nature and extent of their engagement with the quotation may shape the quality of their work. They may also not realise which aspects of their old routines are no longer appropriate (Schwartz et al., 2012). Students in this study were aware of the differences with A-level, and the need to adapt to university writing, but found resources such as assessment criteria opaque. As Holly put it at the beginning of Chapter 5, 'technical jargon' used in assessment criteria, and in examiners' description of writing is not the issue. Rather, students without experience of assessment lack a 'benchmark', or the tacit meanings behind such descriptions, and so cannot fully work out the implications for their own writing. What does 'an engaged and questioning reaction to those unseen provocations' look like in practice?

This chapter addresses how students form task perceptions in the face of such challenges, and how assessing example and peer essays informed task perceptions during the study. From the self-regulated learning perspective that informs the theory of action, task perceptions are both the first phase of writing, and the product of that phase: an interpretation of the question which shapes how writing then unfolds. This chapter examines students' metacognition and actions as they select and begin to engage with questions. It particularly examines how knowledge, prior experiences and emotion shape task perceptions, and how participating in the workshops raises awareness of the importance of the question. The chapter primarily addresses, then, Research Question 3: (How) do students' task perceptions shift over the course of the study? The chapter begins with an overview of task perceptions at group level, before moving on to a focal case, Ben, whose approach to the quotation-question shifted considerably. As described below, Ben represents the wider trend across the group, and his material is particularly rich as he both completed a post-workshops think-aloud and annotated the workshop material more copiously

than others. The other focal case is Chloe. Her case is important because it further highlights the role of emotion and self-efficacy in task perceptions and the self-regulation of writing.

7.2 Developing task perceptions: an overview

Before students start writing, the theory of action posits that they engage in two pre-writing phases: task perceptions and planning. In this study, task perceptions consisted of three main operations: reading the question paper, choosing a question, and interpreting that question. Planning tended to involve linking material to the question, recalling material, developing an argument, and ordering the essay. In many cases, the boundary between the two phases was not distinct, and students cycled between the two. For instance, at Interview 2, Phoenix's think-aloud shows that he started planning two questions in turn in order to inform his decision. Once he had chosen, he linked his ideas to his interpretation of the key words of the question – 'the power of the written word'. He also revisited the quotation in full for further interpretation before beginning to write, with his introduction addressing the quotation from the first line. A further aspect of task perceptions influenced Heather's planning stage – she realised that in the three-hour exam itself, she would have to be careful about dividing up her material between the three essays. This requirement to not repeat material became a standard for her planning, as she rejected points which she might need to use in another essay. Planning, too, can seep into the main writing phase of enactment. For example, at Interview 1, Amber discovered her thesis while writing her introduction, and started again. Task perceptions also shifted during enactment for some participants as the knowledge-constituting process of writing opened up different angles on the quotation (Galbraith, 2009).

In the timed task of one hour, even with ten minutes of extra time given at Interview 1 to compensate for thinking aloud, students did not spend long on task perceptions and planning. It took participants a mean time of 3:56 to read the paper and choose the question at Interview 1 (range = 1:02–9:37; SD =2:44). At Interview 2, this was slightly longer: 4:48 (range = 2:59–7:39; SD = 1:59). Question papers typically contain at least twenty items. Under time pressure, students read strategically. As Felicia put it at Interview 2, 'I read it slower at first, and then I got like two pages in and I thought right, OK, I just need to scan [laughter]'. Others did not read all of the questions, as we will see in Ben's case. As mentioned in Section 6.6 and shown in Appendix N, the most frequent standard code during task perceptions was *knowledge-question match*, and students often immediately linked questions to their topics or authors, or rejected question instantly with a simple 'no' (Holly, Heather) or 'don't like that' (Chloe, Phoenix, Tessa). When questions raised possibilities the response was often 'interesting' (Heather, Megan, Nicole). There are some indications that the quotation itself became an important standard at Interview 2, as Nicole explains:

> '…[it] was a broad question but it had something specific in the quote, because we talked about engaging with the quote and so that was definitely something which I was trying to do. Yeah… yeah, I wanted to be able to... err, I don't know, quote it or at least be engaged with part of, different words it says, and its argument.'

As discussed above, the question and quotation are sometimes in tension: here, Nicole's quotation is 'specific' enough to engage with, and the question is 'broad' and so does not impose a particular interpretation on the quotation. Potential conflicts between the question and quotation will be discussed later.

The time between deciding on a question and starting to write the first line of the introduction was slightly longer: a mean time of 8:03 (range = 4:30–20:48; SD = 5:53) at Interview 1 and 5:13 at

Interview 2 (range = 3:22–11:31; SD = 3:32). This time was predominantly used for planning. However, as the workshops highlighted the importance of the question-quotation to quality, post-workshops data indicates that some participants spent more time interpreting or revisiting the quotation, as we shall see in the section on Ben, next. Results of coding given in Appendix N shows that across the group, uses of *question* as a standard during planning increased from four references at Interview 1 to twelve at Interview 2. As the overview of large-scale adaptations given in Table 12 in Chapter 5 shows, many of the writers who made changes linked to the workshops increased their focus on the question. For Megan, Phoenix and Ben, there is clear textual, and for the latter two, think-aloud data for them deconstructing the quotation during the task perceptions phase. In addition, for some writers, the products of the task perceptions phase – their interpretations of the question – became standards which they used to monitor writing to ensure it remain focussed, and also shaped their topic sentences. As we will see in the case of Chloe, sometimes this interpretation involved narrowing the quotation to make it more manageable.

In the theory of action, the products of task perceptions can be used as standards during writing. Appendix N shows that across the group, the instances where task perceptions were used as standards during enactment increased from nine to seventeen. In other words, participants were monitoring their interpretations of the question-quotation and how their writing aligned during writing. An example of this in practice was Amber, who at Interview 2 commented:

> 'I was realising that I had to like do a bit of legwork to make the quotation work for just that fact. Because I didn't want it to be like [pause] that it's general and rare and extraordinary, because that's not what the quotation was saying at all.'

Amber here is aware of the potential for her essay to shift beyond what she sees as the essence of the quotation – her interpretation of 'what the quotation was saying'. Similarly, Phoenix used his interpretation of the question to monitor the relevance of his 'evaluation' moves:

> 'The only kind of like sticking point was making sure that in my evaluation I made it relevant to the question, not just relevant to a particular concept I had in my head.'

His use of 'sticking point' shows how this is not yet an easily deployed routine, but an intentional, metacognitively monitored act.

At Interview 2, only one writer did not explicitly reference the quotation in her essay: Holly. As described in Section 5.8, a key purpose for Holly in the Interview 2 essay was to develop her argument on a topic, and to structure her material. To prepare for Prelims, Holly continued with a practice developed at A-level, writing and self-assessing essays, beginning with a 'fantastic sort of tutorial-style essay, everything that I have' and graduating towards timed essays responding to generic themes rather than specific question-quotations. This was carried over into her approach in Interview 2, where the think-aloud protocol captures her reading only the question, and not the quotation. The interview confirms this:

> 'When I chose the question, I hadn't really looked at the quotation. I thought that was the one that I want to do, I'd glanced over it and seen err yep, filth and dilapidation, [researcher laughs] that sounds like a pretty good description of the experience. So when I was going through, I actually reread the quotation and thought OK, have I covered it, have I seen where it's come from.'

Holly does check the quotation then, as she writes, but this is about pure task compliance rather than the quotation adding to the quality of the essay. Her emphasis in writing is on rehearsing her knowledge and developing an argument, and this has shaped her task perception to be about just that – not interpreting and checking the meaning of the quotation. The day before Interview 2,

Holly had 'just forced myself to write a timed essay' on a topic with which she was struggling. She wrote on this topic again, and the two essays overlap in the introduction and first two paragraphs; the sequencing of material diverges in the third, where she discovered a new idea and her argument shifts, allowing her to write a fourth paragraph for the first time. To manage this shift in argument, she spent more time on the conclusion. In the interview, Holly expressed some 'concern' that the introduction didn't set up this new 'implication' of the argument, but in the ten minutes spent checking, did not make edits to address this. It seems then, that Holly was writing for herself; there is less concern about clarity than other participants. Her purpose was to work out her arguments, rather than make additional adaptations to a system with which she is happy.

Cross-case comparisons are thus difficult to make as participants hold different goals for writing, as well as specialising in different topics of the period, and answering different questions. In this chapter, I have chosen as focal cases Ben and Chloe for several reasons. Firstly, both opted to complete a think-aloud at Interview 2, which gives greater insights into their decisions and processes. Ben is a representative case for the group's increased emphasis on the quotation at Interview 2. In contrast to Tessa, discussed in the last chapter, and to Chloe, Ben has a strong preference towards planning thoroughly before writing rather than allowing ideas to develop through knowledge constituting and then revising (Torrance et al., 1999; Galbraith and Torrance, 2004). Ben is also an ideal case in terms of data: as well as completing a think-aloud, he made copious annotations and additions during the workshops, and was consistent in his use of coloured pen, allowing the influence of the discussions to be traced. Ben and Chloe's think-aloud protocols were also relatively fluent; as Ericsson and Simon (1993) acknowledged in their seminal text, not everyone takes to the method. For example, Holly tended to narrate behaviour rather than purely verbalise. Finally, emotion emerged as important to writing in this study. This was not a focus of the interview guide, but Ben and Chloe were more open than other participants, and it

is important to highlight the often over-looked role of emotion in assessment, self-regulation and writing.

7.3 Focal case: Ben

Figure 14: Ben's question and introduction at pre-workshops

Ah Nature, framed in fault,
There's comfort then, there's salt;
Nature, bad, base, and blind,
Dearly, thou canst be kind.

(GERARD MANLEY HOPKINS)

Discuss Victorian attitudes to nature.

The **relationship** between man and nature is both complex and diverse. Indeed, the poems of Gerard Manley Hopkins and Elizabeth Gaskell's 'Mary Barton' show Man's ways of conceptualising nature. Nature is the world around us, on a basic level everything not made by man (including the man himself). However even this vague definition is unstable – where does the world of man end and the world of the natural begin, if such a distinction can even be made? For Hopkins, the transcendent link – in a sense of the foundation behind both of these 'worlds' – is God. Poems such as 'Pied beauty' and 'God's Grandeur' Hopkins praises God's presence in the natural world and implores man to look through it properly. Gaskell's 'Mary Barton' depicts, through Alice Wilson, a situation of human dependence on nature. Although the difference may appear in the aspects of nature the two writers explore – it's [sic] religious or its socio-economic importance, what is important to note is that how these two writers explore a changing and sometimes detached **relationship** between Nature and Man.

In the pre-intervention writing session in February, Ben chose to answer on a question on nature, which accompanied an extract from a Gerard Manley Hopkins poem. Above, Figure 14 gives the question, and his opening paragraph. In taking its focus as 'the relationship between man and nature', Ben's introduction appears to have some of the characteristics of the examiners' 'downloaded' essay. The quotation could reflect Hopkins's complex relationship with nature, but Ben does not refer to the quotation in his introduction, and so this connection is not explicitly made. The question frames the quotation as being about 'attitudes', but Ben doesn't refer to this

keyword. His introduction is clearly about nature, but by not referring to the Hopkins quotation or question, his answer loses relevance.

The interview and think-aloud data give insights into the task perceptions and planning that shape Ben's introduction. He comes to the session wanting to write about Gerard Manley Hopkins, and this goal is clear as he reads the first question on the paper in full, then rejects it: 'still want something to do with nature given that umm focussing on Hopkins, let's look for one that looks, look for one that... that discusses nature or religion.' After this, he scans the paper, verbalising only questions, going through seven items in rapid succession until:

> 'Ah! Discuss the Victorian attitudes to nature. The quote is um Gerard Manley Hopkins who I want to focus on, so that helps. That's definitely useful to be able to like link in the quote to what I'm going to say in my essay. He says: "Ah Nature, framed in fault, / There's comfort then, there's salt;" / Salt, comfort, um why the juxtaposition? Nature, bad, base. Base basic or base atavistic? "And blind, / Dearly thou canst be kind." Can't be kind. Why is he, does he mean to suggest that nature's unable to be kind?'

He chooses this question 1 minute 40 seconds into reading the paper. That the quotation is Hopkins 'helps'; it is 'useful to be able to like link' it to his argument, but the primary standard for selecting the question seems to be the opportunity to write about the topic he has prepared. Ben begins to interpret the meaning of the quotation, but doesn't go further than the extract above, instead returning to the question, 'discuss Victorian attitudes to nature. Discuss Victorian attitudes to nature". Umm. Victorian attitudes. Umm. I could compare it a bit to Elizabeth Gaskell, if I could remember a quote.' At this point, he seems to perceive the task as comparing Hopkins and Gaskell on nature. His focus switches away from the question and quotation and entirely onto his material. For the next 12 minutes, Ben's think-aloud protocol shows him recalling material from a mindmap he had prepared for college collections, initially retrieving information about Hopkins and nature:

> I'm like drawing a mindmap, Hopkins and then nature, one of the things that comes out of it, nature, umm, then, there's the four things I remembered, it was, what was it? There were four, and one of them was 'fresh-', umm, what was the quote about?

From here, Ben recalls points and quotes, noting them in mindmap form on his written plan. As he recalls, he questions both his own knowledge and the meanings. In the interview, he states that he sees a good exam essay as 'probing… always questioning'. After 12 minutes, he revisits the quotation, and this questioning stance can be seen in action:

> I think it's important to go back to the quote, what does the quote say? "Ah Nature, framed in fault, / There's comfort then, there's salt"; salt, salt of the earth? Is that what it's saying, this idea of like the root of the earth, working class, kind of rural affinity to nature, or the rural working class's affinity to nature? "Nature, bad, base, and blind, / Dearly thou canst be kind." Why is he saying that? Is it against…? Maybe if it's umm if you view it as, view nature as a part, aspect of- of God, then nature becomes a is it a metonym, or is it a metaphor?

Although Ben sees returning to the quotation as 'important', he doesn't seem to have a distinct purpose for doing so. He seems to still be puzzling out its meaning, and doing so sparks further ideas related to his material. Shortly after, he begins to focus on the 'relationship' between man and nature as the comparison point between Hopkins and Gaskell, and this features in his introduction. Towards the end of the protocol, he checks the time and shifts from recall to planning his opening and structure for the essay:

> I've got 15 minutes into my planning time, how am I going to start it? Start off with nature and talk about, so it's **attitudes** to nature, nature as-, I think I'll start off with the idea of nature as religious, start off with an inscape quote and talk about beauty of it, nature beautiful, nature religious, start off with nature as religious then intersect it with ideas from the nature as beauty point and then umm compare it to the Gaskell, nature is with a purpose, and that should be enough. OK. I'll start writing. (my emphasis)

This is the first time he revisits the question, 'attitudes to nature', since choosing it. This seems to be a quick validation check, rather than a further interpretation of the question. He doesn't revisit the quotation again. This last minute of pre-writing time is spent developing a sequence for the material: religious – beautiful – purpose – and then he considers 'that should be enough.' However, while writing the introduction, a new idea arises in the interaction between his task perception of the question being about nature, and his standards for an introduction, which seem to include defining key terms: "I really need to explain what nature is." But then what is nature?' His attempt to define nature can be seen in his introduction (Figure 14, green text).

Task perceptions are formed in the interaction between the set task and the writer, and what they bring to the task (Winne and Hadwin, 1998; Flower and Hayes, 1980). In Ben's think-aloud, his material is the dominant player in this interaction, and he only briefly unpacks the question and quotation before moving into a planning phase. The task perceptions phase itself is very brief: most of his pre-writing is planning, or recalling and then organising material. The preponderance of recall is explained by his exam preparation strategy. Rather than learning quotations from his texts, as other students do, he memorises quotations with ideas. These are in the form of 'massive' mindmaps made up of 'little pockets of information' gathered from tutorial essays. He describes the mindmaps as 'both quite fluid then quite fixed': the 'little blocks' are fixed, but can be 'put together in different ways, depending on what the question is'. In the think-aloud protocol, he recalls the blocks, and then, in the extract above, sequences them: '…start off with nature as religious then intersect it with ideas from the nature as beauty point'.

Ben first developed this mindmap strategy for his GCSE exams. He acknowledges that with increased reading and challenge at degree level, 'they're quite big so it takes quite a long time and it hurts, the revision is just painful because it just hurts your head trying to remember all the stuff'.

He considers that while 'it could be more interesting' to deviate from the plan, learning and using mindmaps is worthwhile.

> 'I want to make sure I stick to that because then I can know how long each idea takes to write approximately and you can kind of get it out, and it means you can do it regardless of how really you're feeling on the day. […] The whole point of the exam is to like, at least in my opinion, is to, placate as much of the risk of doing it like, messing it up, as possible.

The mindmaps are, then, a risk management strategy, guaranteeing that he will be able to get his knowledge onto the page within the time limit despite any exam anxiety, and so avoid 'messing it up'. Having been through high-stakes GCSE and A-level exams, performance is clearly important to Ben. His conception of the Prelims task, though, is that it is not at all prescriptive. In response to the interview item, 'what is the question asking you to do?', Ben describes it as 'open-ended':

> 'They'd accept any, like going from the extreme of like just a pure close analysis text if it's done well. Or a complete [pause] completely theoretic […] you can do a theory essay on the like literary theory'

Ben's response reflects the course's valuing of diverse approaches to literature. Interestingly, this description matches his description of a tutorial essay: he describes one tutor as preferring theory-based essays, while others expect more close analysis. As in Alder's (2018) interview research with students of English, his tutors hold and communicate different expectations for essays. When prompted to reflect specifically on the question he answered, Ben considers that only two things are fixed:

> 'The two things it has to be is Victorian writers, obviously **writers from the period and nature**, erm, [pause] but yeah I think, with, with regards to like the particular question it's, **as long as those two boxes are ticked it's so open-ended it's fine'** (my emphasis).

This post-writing task perception both aligns with and explains the think-aloud data, and his emphasis in planning on retrieving ideas about nature, and his two writers of the period (Hopkins and Gaskell). Following these task perceptions through – ticking 'those two boxes' – leads him to assume he has achieved task compliance. This misconception perhaps originates from his experience of writing tutorial essays, in which students tend to have great latitude, as long as they write on that week's topic and/or authors. In a further follow-up question, I asked, 'what do you think is the function of the quotation?'

> 'I think they are, they're [pause] as [pause] kind of a prompt. [pause] […] because obviously you've such an open-ended question, it needs something […] **it's a way of helping those who perhaps haven't got like a set essay, not a set essay but it's a way of focusing it [pause] without [pause] without like imposing that focus on the, on the question in general.** So you can either… just look at… this is what I did, "discuss Victorian attitudes to nature" and keep it completely unrelated to the question, not the question the quotation. Or, **if you haven't got anything, like you haven't got your own idea of an argument, sometimes you can use the quotation and that's almost a prompt in itself.** I don't think there's a problem in just doing it on the quotation, no on the question without the quotation. At least I hope not [laughs]'

Ben comes across as unsure in this extract, and the self-correction of 'not a set essay' is perhaps telling. Such uncertainty is understandable – he had only written one Prelims-style collections essay before the meeting. Nonetheless, his response matches the sense in his think-aloud protocol of the quotation as being optional - he acknowledges that he chose to 'keep' his essay 'unrelated' to the quotation. He sees the quotation as a way of focussing the broad questions, and also as helpful for those who don't have 'your own idea of an argument'. Quotation and argument are separate, then, and arguments are conceived prior to the exam, in tutorial essays, and in mindmaps.

7.4 During workshops: 'DON'T IGNORE THE QUOTATION'

During the workshops, Ben's understanding of the role of the question and quotation develops, as evident in data such as his annotations and responses to example essays, and the additions he makes in different coloured pen during discussions.

In his response to the first example essay, Ben initially comments that:

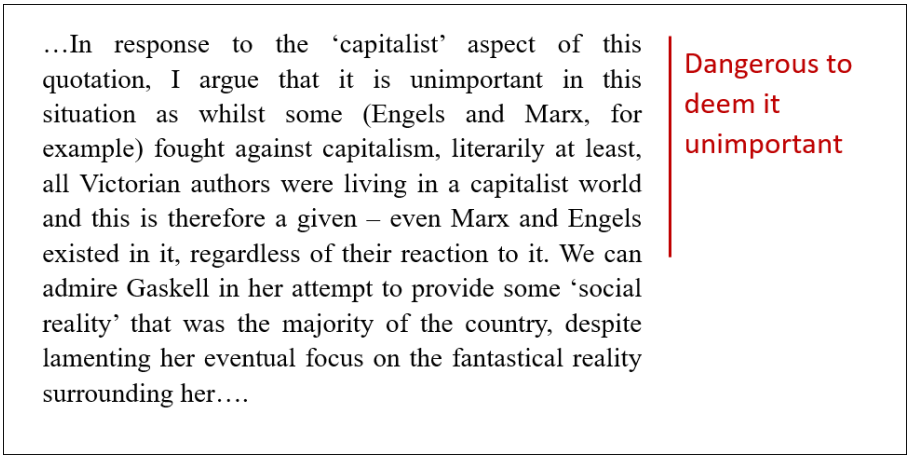
> 'I like how they refer back to the quotation in the conclusion, even if the response is slightly garbled. Shows an awareness of the question and a desire to engage with it.'

Here, it seems Ben sees the quotation as an optional bonus – he credits the writer's attempt. However, during the discussion, different views are raised, and he makes an addition in a different coloured pen (Figure 15, below). Thus Ben realises that engaging with the quotation only to dismiss it is 'dangerous'. As Amber, who was in Ben's workshop group, wrote in her response to the essay and also raised in the discussion, 'the examiners wouldn't have used a quotation if parts of it were to be "unimportant" as an afterthought in the conclusion.' Having assessed this and two other essays, and developed a set of criteria, Ben reports in end of Workshop 1 reflections:

> 'For me, and my writing, the most important [criterion] is engaging with the Q and the quotation. By setting specific parameters in relation to the question, an essay's argument can become more relevant & more convincing.'

Engaging with the quotation and question offers more than task compliance, then: it can enhance quality of argument. In Workshop 2, students assessed Example Essay 4 before writing their own. During the discussion, Ben again makes further addition to his notes: 'DON'T IGNORE THE QUESTION – use it to focus arguments'. As well as this self-instruction, he also adds to his annotations on the introduction, drawing a line to the writer's thesis statement ('it is my

Figure 15: Ben's addition to the conclusion of Example Essay 1



contention that…'): 'use this type of opening argument'. With reference to later points in the introduction, he adds: 'make clear to examiners – the points are there but not flagged up' (his emphasis). These additions are again phrased as self-instructions, using imperatives.

In his workshop essay, Ben opens his essay by directly addressing the quotation: 'Ramazani may be correct to argue that 'imperial imposition' is 'transnational', but unlike 'global dialogue' it limits the potential for multiple narratives.' He connects this to the authors he wants to write about. He initially disagrees with the quotation, but unlike the first example essay, uses evidence to do so – his interpretation of an author's stance. However, in Workshop 3, on re-reading his own essay after assessing peer essays, Ben realises that 'I don't have a clear thesis in my introduction' (his emphasis). To address, this, he considers that 'I would spend more time planning – get a clear and developed thesis clear in my mind before I start writing.' There was no evidence of Ben developing a thesis in his Interview 1 essay, so this reflects a different purpose for pre-writing time. In feedback, which is discussed in more detail in Section 8.7, Ben's reviewers also raise the problems he noticed. For example, Chloe suggests that 'at the end of your introduction include a sentence that makes your own stance really clear such as 'I argue…' as this will help your essay be easier to follow as a reader.'

7.5 Post-workshops

In Ben's post-workshops essay, two adaptations are evident in the task perceptions phase of his writing, and in both interview and think-aloud data. In the interview, Ben perceives that his writing has developed over the course of the study: 'From the one from the very beginning that I did with you [laughing] it's completely – not completely changed but I engage a lot more with the… like all the aspects of the question.' As well as focussing on the quotation, Ben formulates a thesis statement. He connects this to his workshop learning as follows:

> 'I wouldn't now remember one essay and think 'I have to write like that person wrote in that essay' but you kind of get a sense of, that good essays have this kind of… umm, what was it, like control over the subject matter, and they have a strong argument, and it's very clear. I think clarity is one thing I really – and also, yeah, clarity is one thing I really got from the workshops, especially in the introduction, I set out the thesis.'

'Setting out the thesis' is, then, a way of working towards *clarity* and *argument*, the wider importance of which was addressed in Chapter 6, on students' personal standards. In formulating his thesis, he doesn't 'copy' the strategy he picks out in Example Essay 4 ('it is my contention that') or use the 'I argue' format that Chloe suggests in feedback. However, the last sentence of his introduction, given in blue in Figure 16, does outline his argument.

Figure 16: Ben's question and introduction at post-workshops

Each for himself is still the rule:
We learn it when we go to school –
The devil take the hindmost, O!
(A.H. CLOUGH)
How do Victorian moralists regard individualism?

The **moralist** perspective on **individualism** functions on the assumption of a single, integral self. **In Clough's quotation,** 'each for himself' uses the singular reflexive pronoun, the individual is conceived as [illegible] and integral. However, by the late-19th century, writers such as Robert Louis Stevenson, Henry James and Oscar Wilde were supporting an increasing vocal portion that **questioned this assumption of an integral self that underlines moralist thinking. Their texts, complemented by the later essay 'The Uncanny' by Sigmund Freud, show the disintegration of the notion of an integral self, a disintegration that destabilises the assumptions of late Victorian society, specifically with regards to savagery and civilisation, too.**

At Interview 2, the think-aloud protocol shows that Ben structures his time differently. After deciding on the question, he spends 11 and a half minutes forming task perceptions and planning. Whereas at pre-workshops, the distinct task perceptions phase was fleeting, Ben spends four minutes interpreting the question to arrive at an initial thesis statement. He then spends four and a half minutes recalling and organising material, before using the last three minutes to deliberate on the terms he needs to use for consistency and refine the content of the introduction and thesis statement. He returns to the question and/or quotation on four occasions. At Interview 2, Ben is still 'learning' his mindmaps for revision, to which he has added 'more texts to give it more scope', and also some theory. However, as the break-down shows, the mindmap content is less prominent – unlike before, he doesn't draw one out. When I ask Ben what the question is asking him to do, he explains that 'they're trying to get you to show how well you can fit your knowledge to a question'. As the analysis in the next paragraph shows, this task perception is reflected in his

think-aloud protocol. It also echoes and aligns with that in the 2014 examiners' report that opened the chapter.

As with Interview 1, Ben decides on a question fairly quickly – in three and a half minutes. Again, he does not read the whole paper. When he first encounters the item he later chooses, he reads through the quotation once, and even before verbalising the question, re-reads it and starts to interpret it, using close reading:

> 'Each for himself if still the rule', like self-centredness, umm, self, each for himself, the fact that that implies a stable self. 'We learn it when we go to school'. Umm, society [writing] umm, society is umm like sort of like society is… forcing it, or… pressurising, or make, yeah, pressurising people into, people into certain ways, into certain ways of thinking. [writing]. 'The devil take the hindmost, the hindmost, O! Each for himself is still the rule, We learn it when we go to school, The devil take the hindmost, O!' 'How do Victorian moralists regard individualism?' Moralists [writing] [unintelligible] talk about moralists or talk about, deconstruct what moralist is, umm, [writing] and say it's not moralism.

There is a sticking point: Ben identifies the term 'moralists' in the question, but as became clear in the interview later, had not encountered moralists in his teaching. He immediately seems to have a possible tactic for dealing with this, though: 'deconstruct what a moralist is, umm, [writing] and say it's not moralism.' When he decides on this question shortly after, he again re-reads the question in full, and reflects on the 'implication' of the quotation.

> 'I think the implication of the question is that [writing] the individual is stable umm [writing] 'How do Victorian moralists regard?' I need to have one bit where I actually address the question and go on from it.'

The 'moralists' are still a problem which he has to resolve. A clear standard for his task perceptions phase of writing is setting up 'one bit where I actually address the question', and he perceives 'moralists' to be a key part of that question. He puzzles over 'what does it mean by moralists?',

and then arrives at a possible way forward: 'I could put in a thesis saying what they do and say how it contests other people's'. He starts to formulate his thesis, which reflects the final sentence of the introduction that he writes later:

> 'they regard it… [pause] do they need the self to [writing] to [pause] umm uphold society's values, not values, uphold the structures. Structures? Narratives of society. Umm need a stable self because of [writing] threat of primitivism'

Once Ben has a tentative thesis, he then begins to recall and organise material, returning to the question and quotation more frequently than at Interview 1. Task perceptions and planning are less clearly demarcated; it is a recursive process. Despite moving towards what examiners value in task perceptions – 'think[ing] out the new question on the spot', as the 1935 report put it, Ben acknowledges that his essay may not achieve task compliance. When I ask him about the question in the interview, he reveals he may not have used this approach in the exam itself:

> Well, [laughing] obviously the purpose of it is to discuss individualism or identity but I think even that, as individualism's an abstract noun, you can just gloss it, not in any way you like but you can inflect it or move it towards what you want, so I was doing… like individual being a stable self which is individualism in a sense, umm how do Victorian moralists regard it? Like I wasn't sure, I - maybe in a Prelims I wouldn't have done this, it was a bit of a high risk not talking about moralists so openly, but like I'm not sure there's a specific set of writers that are the Victorian moralists.

Ben's task perception has shifted a little away from the wording of the question. The question, shown in Figure 16 (above), is 'How do Victorian moralists regard individualism?'. Yet Ben sees it as 'discuss individualism **or identity**' (my emphasis). Identity is his interpretation of the quotation, not the question stem. In his task perceptions, Ben was using a strategy suggested by a college tutor: to 'gloss' the quotation, and if necessary, 'move it towards what you want'. The bold text in Figure 16 shows how he does this in his introduction. Although Ben doesn't pick up the shift from

individualism to identity, he does find that the moralists are a sticking point, and acknowledges that 'it was a bit of a high risk'. The research situation here comes the fore: Ben's goal for the session was different to the exam itself: he wanted particularly to write on this material as part of his preparation. He is at least partially aware of how he may have not achieved task compliance. The think-aloud protocol shows that in contrast to the pre-intervention essay, the essay process was not 'downloading' material but rather forming task perceptions and a thesis first. 'Moving it towards what you want' is perhaps not an ideal strategy, but it is a valid one. In a real examination situation, students could face not being able to find questions which suit their material, given that the questions must sample the whole period of 1832-1910, and there is no set curriculum or core of knowledge to be assumed. In Ben's case, he can inflect the quotation, but the question here serves to complicate rather than scaffold his engagement: he is unsure whether there is a distinct group of Victorian moralists, and so the question serves to narrow the quotation. The wider implications of this will be further examined in the Discussion chapter. Next, I examine the role of affect and self-efficacy in task perceptions, referring to Chloe's use of the quotation.

7.6: Chloe's task perceptions pre-workshops

Although the main focus of the study is on the metacognitive aspects of self-regulating writing, affect, emotion and self-efficacy beliefs play an important part, particularly given the study's context. For instance, Ben uses mindmaps, even though they 'hurt' to memorise because they ensure he can 'get [knowledge] out… regardless of how really you're feeling on the day' of the exam. This is an emotion regulation (Wester and Hadwin, 2014) as well as a cognitive, or knowledge-transforming strategy, then (Galbraith, 2009). In other instances, affect, emotion and self-efficacy shape the decisions students make during writing, and Chloe is used here as a focal

case as these are particularly prominent in her data. Like many other participants, Chloe perceives Prelims writing as 'stressful' because 'you know if you had a longer time to do the question you would do it really well'. Time is an immediate threat to self-efficacy as it creates the expectation that writing will not meet the usual standards – those of a tutorial essay.

Chloe's writing at Interview 1 is influenced by her experience of writing college collections, which she describes as 'horrific'. As she 'was just really, really, really nervous', she didn't have time to write conclusions. In her present essay, she is 'conscious' of the need to finish, and leaves a paragraph in order to follow the idea she has for the conclusion. She reports that this is reassuring: 'I feel like you're making the most of your time that way, even if you're not [laughs]'. In her pre-workshops essay, Chloe also adopts as standards the tutor feedback she received on her collections on the importance of argument and answering the question. This influences her choice of question. As with all participants, Chloe's key standard for choosing a question is a good match with her knowledge, and she rejects any where she isn't 'alerted to something straight away'. Like others, her think-aloud protocol shows her rejecting questions which are 'too vague'. She explains in the interview:

> 'I think I am quite easily scared by questions that give you a quotation and say 'discuss any aspect of this quotation' [laughter] – because I think I'm just automatically like I wouldn't know which bit to take on and… you don't want to pick the wrong bit; you don't know whether you should discuss all of it and so I would always get a bit terrified of those.'

Chloe expresses low self-efficacy for handling this type of question: there is too much risk of 'pick[ing] the wrong bit'. In order to ensure she does answer the question, she chooses one which she describes as 'quite a simple question in comparison to some of the others'. Unlike the more open questions mentioned above, this does not explicitly require her to engage with the quotation:

> 'I'm never sure whether to refer to the quote in the question and in this case I decided not to, which probably was not a good idea but umm I find when you're looking at too many things, it can often be a bit confusing, you don't want to sort of get through an argument with the quotation if you think it's going to sort of impinge on it in any way.'

Like Ben, then, Chloe at pre-workshops sees the question and argument as separate. She is acting on her tutor's feedback on argument, and so sees using 'the quote' as risky - potentially 'imping[ing]'. The think-aloud protocol shows her reading the quotation and linking it to knowledge, but to answer the question, she focuses on a key word: 'representation'. Her introduction, in Figure 17, shows a strong focus on representation ('presented', 'deemed', 'evident' – in blue) as well as money and also morality (in bold).

Figure 17: Chloe's question and introduction at pre-workshops

'I can do a good deed with my cash. My thousand a year is not merely a matter of dirty bank-notes and jaundiced guineas (let me speak respectfully of both though, for I adore them); but, it may be, health to the drooping, strength to the weak, consolation to the sad' (CHARLOTTE BRONTE).
Discuss the representation of money AND/OR morality in writing of the period.

Money and morality are inextricably linked in the literature of the Victorian Era, and the relationship between the two is often **presented** as antithetic. An abundance of **money**, if it is not used for utilitarian purposes, **is often deemed immoral, indicating a selfish, greedy character.** Conversely, **a lack of money,** not necessarily complete poverty, **is often a way of creating a humble, inherently good character**. Whilst this relationship is by no means definitive it is nevertheless **evident** in Dickens' Hard Times, Eliot's Silas Marner and Daniel Deronda and Charlotte Brontë's Jane Eyre.

7.6: Chloe's task perceptions post-workshops

At Interview 2, Chloe makes fewer adaptations to her task perceptions than other participants. This is perhaps because in the workshop data, Chloe comments less than other participants on the *question*. The most prominent connection between the question and her own writing is in Workshop 3, after commenting on peer essays. She notices in others' writing 'my own tendency to miss the point of the question and go off on my own ideas'. Instead of channelling her adaptations towards the question, however, Chloe focuses on finishing her essay for the first time, achieving this by reducing the number of texts. She also reports focussing more on the argument and outlining the structure in the introduction.

As at Interview 1, Chloe chooses a question which 'wasn't basic, but there wasn't anything too sort of much to be inferred from it'. She does, however, report trying to take more account of the quotation. She writes down the question on her plan, which is kept in front of her throughout writing, and underlines parts 'that I knew I could talk back to, just to make sure that I didn't ignore the quotation'. Her standard here seems to be to ensure task compliance, rather than use the quotation to enrich her argument. Chloe's underlining can be seen on her question paper: she seems to break the quotation down into '<u>We none of us spoke of money</u>, because that subject savoured of commerce and trade, and though some might be poor, <u>we were all aristocratic.</u>' This re-shaping of the quotation is evident in her essay's emphasis on class and capitalism. Unlike the pre- and during-workshops essays, she refers to the quotation directly: once, in paragraph 2, introducing a point about individualism as seen in Figure 18. However, her reduction of the quotation to the underlined parts 'that I knew I could talk back to' leads her to misrepresent it. Gaskell's omitted reference to 'some might be poor' seems to refer to some of 'us': the aristocratic, rather than to those who are of other classes, and therefore doesn't imply social

complicity in poverty. Her reduction is understandable as the quotation itself seems more immediately suited to the first option, (a) class.

Indeed, in the interview, Chloe sees the question as '*just* how Victorian literature portrayed commerce and capitalism, you can't really go any more to that, [laughs]' (my emphasis). She doesn't see the quotation or question as requiring further unpacking or deliberation, and this is reflected in the think-aloud protocol. After choosing the question, she verbalises for four minutes before starting to write the introduction. Initial task perceptions seem no more than rereading the quotation, and she goes straight into possible points, for example: 'what other novels can I do? If I change that to class, I could add Wilde in.' Showing range is still an important standard, even though she reduces the number of texts she handles in order to finish. There is one review of the truncated quotation: ''We none of us spoke of money, we were all aristocratic.' Could link the quote to, I did social complicity [writing]'. This link is the one enacted in paragraph two, in blue in Figure 18 (below).

Chloe's introduction uses the key term 'capitalism' repeatedly (in green), and in the interview she acknowledges that she has narrowed the focus:

> 'I don't think I addressed 'commerce' as much as I should have done, it was 'commerce and capitalism', but I thought they both come under quite an umbrella term, so that should be OK.'

Indeed, when prompted to talk about strengths, Chloe judges that 'it's quite clear, I think I've stuck to the question…'. So, for her, she has managed to achieve task compliance, her standards for which seem to be using keywords throughout and referencing the quotation. Although Chloe doesn't mention again in this interview that she sees quotation-questions as 'scary', her main goal

Figure 18: Chloe's question and introduction at post-workshops

'We none of us spoke of money, because that subject savoured of commerce and trade, and though some might be poor, we were all aristocratic' (ELIZABETH GASKELL). How does Victorian literature treat EITHER (a) class OR (b) social commitment OR (c) commerce and capitalism?

Victorian Literature **treats capitalism**, on the whole, as one of many social problems, particularly when its foundations and conditions are taken to the excess. George Bernard Shaw and Charles Dickens are examples of two Victorian authors whose works critique **capitalist forces**, using microcosmic examples in order to gesture to the issues in society at large. Both focusing on social problems in their prose and critical reading. In 'Widowers' Houses' (1892), 'Mrs Warren's Profession' (1893), Bernard Shaw reveals a society that is tainted by **capitalist motives**, and complicit with its ills. Similarly, in Hard Times, (1854) Dickens uses Louisa Gradgrind and Stephen Blackpool to elucidate the different, yet equally detrimental implications of **a capitalist fate.**

[...]

His decision further on in the play, to accept this money, thus indicates that he has neutralised himself to the exploitation of the poor, its source. A critique by Shaw on social complicity, **as Gaskell highlights in the quotation "we none of us spoke of money... we were all aristocratic". Capitalism** appears to breed individualism, as long as the individual is financially stable the concept of money can be ignored, and therefore, so can social responsibility, breeding immortality.

was to finish in time, which she achieves, and with a longer essay than before. As reported in Section 5.6, experiencing the frustration of unfinished essays from a reader's perspective, she focuses on 'getting things done'. However, this focus on finishing doesn't leave space for her to take a risk. As discussed in Chapter 6 and in Chapter 8, participants in this study did not have much time to adapt, and so prioritised particular areas. There is perhaps less motivation to deal with aspects of writing that are 'scary' if other vital personal standards can be met. For Chloe, finishing in time is important, and she has complied with the task by 'talking back' to the quotation, even if examiners would value a more elaborated interpretation.

7.7 Chapter conclusion

This chapter has highlighted the challenges students face during the task perceptions phase of writing, and the influence of prior writing experiences. During the workshops, *question* became the most prominent criterion and Ben's case shows how seeing for himself the importance of the quotation and a clear thesis influenced his approach in the post-workshops essay. For Chloe, the experience of writing a collections essay, and the feedback she received was also influential. In both cases, emotion played a role in writers' task perceptions. The need to mitigate the risk of 'messing it up' (Ben) or 'pick[ing] the wrong bit' (Chloe) shaped decisions in the opening minutes of writing.

The chapter has also shown shifts in the strategies used to develop task perceptions and address the question in the introduction. For Chloe and Ben, they perhaps do not use the quotations 'to significant effect' or as 'provocations', as the examiners desired in the 2015 report. But they show increased engagement with the question and quotation, and in adapting their knowledge to the task. Ben's case shows how students' development may not be visible in their written texts. He adapted his approach to pre-writing, spending more time on interpreting the question and quotation, and also formulating a thesis statement. The products of this are visible in his written introduction. However, although he spent much less time on recall and referred to the question and quotation during planning, and addressed it in the introduction, his essay may not achieve task compliance. This is because he didn't manage to find a question that matched his knowledge more appropriately. Likewise, at post-workshops Chloe ensured that she 'talked back' to the quotation. Yet the quotation itself was perhaps not as well suited to the question she chose as to the other options, and she condenses it, making it more manageable.

Both Ben and Chloe, then, tried to address the tension between what they wanted to write about – the literature they enjoyed and had worked hard to prepare – and the demands of the question-quotation itself. However, this tension is to an extent unresolvable. The examination papers each sample a century or so of literature. Even given twenty or more questions to choose from, the choice of question, or the nature of the quotation may make it difficult for writers to use their knowledge without resorting to 'moving [the question] towards what you want'. Thus it can be hard to go beyond task compliance and enter into genuine dialogue with the question. This will be discussed further in the Discussion chapter.

Chapter 8

Feedback, or 'new ideas about things that I was already aware of'

8.1 Introduction

Feedback is widely conceptualised as offering next steps in learning, and is often assumed to provide new information about performance to the recipient. However, according to this study's theory of action, feedback is not necessarily new information. As seen in Chapter 7, students make evaluations throughout writing according to their personal, self-set standards, and experience of writing adds to their metacognitive knowledge. Reflections tasks written at the end of each workshop also prompted evaluations and so served to raise awareness. Feedback is filtered through these evaluations, experiences and metacognitive knowledge: students do not come to it fresh (Butler and Winne, 1995). Heather's perspectives in Interview 2 on the feedback she received reflect this:

> 'I thought it [the peer feedback] was really good because they all made really similar points and it was both things I understood about my writing and I was like, "Oh, yeah, you're right about that. You only read one of my essays, but you're right" and also things that I didn't really realise that, like, slightly kind of like deconstructed… umm…. what I did. And so I was sort of like, "Oh, so that's why I do it" or "That's how I can not do it," so it was kind of like new ideas about things that I was already aware of, see what I mean? So I think they understood my writing very well from reading just one essay.'

Heather's feedback both aligns with and enriches her own metacognitive knowledge about writing. Rather than offering her purely new information, her reviewers 'deconstruct' her writing, offering back 'new ideas about things I was already aware of'. This is echoed across the case study dataset. For example, Chloe finds that the feedback 'clarified things', giving her insights into 'issues' of which she had already been aware but couldn't quite 'pinpoint' for herself.

This chapter examines how peer feedback linked into the adaptations that students made, addressing RQ4: (How) does peer feedback contribute to students' self-regulation of writing? As summarised in Chapter 5, eight of the ten case study participants made adaptations to writing which could be traced or attributed to workshop learning. For all of these writers, feedback played some role in adaptations, but not necessarily in the straightforward sense of providing suggestions that were then directly enacted. Students received extensive feedback – the mean length, from all reviewers combined, was 496 words (SD = 90.97). As discussed below, this feedback was rich in solutions. However, in only two cases did students directly implement suggestions made by reviewers in their Interview 2 essay. This low uptake was partially due to limited time to prepare for exams; for example, Heather did not consider she was 'at the writing stage' of revision yet, and so couldn't implement workshop learning as she'd like. Equally, the time constraints of the exam situation itself also limited what participants felt it was feasible to try out. In other studies examining the extent to which recipients implement feedback, participants revised the original task, maximising the potential for transfer; even then the proportion enacted was only a third (van den Berg et al., 2006; Patchan et al., 2016) or half of comments (Walker, 2015). In addition, the framing of feedback in Workshop 4 could have contributed to how students decided to use it. When participants received feedback, they were told to consider it critically, and to decide what they wanted to enact. This was partly because of the ethos of prompting reflection and self-regulated learning, and for ethical reasons, to soften the impact of the very small number of reviews about which I had concerns (see Section 8.7, below).

Nonetheless, data from the workshops evaluation questionnaire indicates that students found value in peer feedback. In response to the question 'which activity did you find most useful?', only six participants selected 'reviewing example and peer essays' alone. Thirteen participants selected 'I found both equally useful', with the remaining student opting for 'receiving feedback from

Table 19: Functions and uses of feedback

Function or use of feedback	Adaptations made by case study writers
Direct implementation of a peer suggestion	Megan *(signposting*); Chloe *(signposting)*
Confirmed own judgement and priorities for adaptation	Nicole *(structure);* Felicia *(question; introduction);* Ben *(clarity; argument)*
Feedback on success of adaptation made in W2 that is then continued in Interview 2	Tessa *(introduction);* Heather *(introduction)*
Raised awareness of importance of an aspect; begins to work towards in Interview 2	Phoenix *(conclusion);* Chloe *(development)*

peers'. Significantly, case study analysis of participants' workshop reflections, interview data and essays shows that feedback served several functions beyond implementing suggestions, as Table 19 above shows. Related criteria codes are given in brackets. For some writers, feedback confirmed the extent to which Workshop 2 adaptations had worked, as we saw for Tessa's introduction in Chapter 6. For others, feedback confirmed students' own views about what they should work on next. For example, Nicole decided to use shorter paragraphs in order to move on with her essay more quickly and finish. In Interview 2, she attributes this to a reviewer suggestion, but in fact had already considered the strategy in her Workshop 3 reflections, when she re-read her own essay after giving peer feedback. Felicia reports not having had time to implement 'specific' feedback, but her adaptations – a tighter focus on the question, and spending more time on the introduction – align with peer suggestions. She sees feedback as 'highlight[ing] stuff that I kind of already knew but you often need it pointing out to you, to make it more explicit, you're like, 'oh yeah!'' (Interview 2). For example, in Interview 1 she is aware of the risk of 'deviating' from the question, and in Workshop 1, the example essays highlight 'what a difference effectively addressing the quote/question makes [to] the quality of the essay'. Connecting this to her own

writing, she reflects that 'I tend to only answer certain aspects of quotations and I often don't engage with the quote fully.' Her feedback, then, further 'highlights' issues, but doesn't necessarily bring brand new knowledge. In other cases, feedback served to heighten awareness of the importance of certain aspects of writing that the student wasn't already working on, as discussed later in the case of Phoenix, who was somewhat persuaded of the need for a conclusion but didn't manage to get to one in Interview 2. In this case, he made another adaptation which seemed to be a prerequisite to the conclusion. In such a time-limited task, participants had to juggle goals and standards.

This chapter, then, examines how, to use Ben's term, students 'amalgamated' feedback with their own judgements, and made adaptations that were guided by their standards, and were feasible in the time. To examine the characteristics of the feedback, the chapter opens by analysing which criteria students drew on in their feedback, and which types of comments they used. This overview complements and extends the findings on criteria presented in Chapter 5, addressing RQ1. I then consider the quality of peer feedback from students' perspectives, comparing feedback which was rated as most helpful with the wider dataset. I examine the role of solutions and praise in writers' adaptations, using Megan's adaptations as an example. The discussion then moves on to consider which types of comments students found unhelpful, and how Ben, who received 'awful' feedback from one reviewer responded to the feedback, 'amalgamated' his reviewers' perspectives with his own. I then discuss two cases where students acknowledged but chose not to act on compelling feedback: Phoenix and Holly. Finally, to address the role of feedback in self-efficacy (RQ5), I draw on instances where feedback provided a 'boost', to use Nicole's term, specifically to self-efficacy for writing, and persuaded participants that 'actually, I can do this', as Felicia put it.

8.2 What do students comment on when giving feedback?

The mean length of feedback that students received, that is from all reviewers combined, was 496 words (SD = 90.97). The mean word count of individual reviews was 158 words (SD = 53.9 words). After receiving feedback in Workshop 4, participants wrote reflections, with a mean word count of 298 words. 17 students also annotated their feedback. The mean word count of the annotations was 84 words (SD = 56 words).

Table 20 gives the frequency of references to criteria codes in two sub-sets of data: the annotations students made on peer essays and the end comments shared with the writer. The table is organised by frequency of references to code. Students' annotations on peer essays yielded 751 references to criteria, and coding of end comments resulted in 1059 references to criteria. As with the wider dataset, references could be coded at more than one criterion, which accounts for the high frequencies of references.

Question is the most frequent criteria code, as it is across all other types of data. The proportion of references coded as *question* is higher in the annotations data set than in the end comments, accounting for 18% of the references as opposed to 12% in the latter data. This indicates that that students looked for engagement with the question in multiple instances in the essays they read, rather than seeing it as a one-off criterion. Interestingly, however, *question* is less dominant than in other data sources: the distance between question and clarity is just 50 references. This perhaps reflects firstly recognition of how multiple criteria work together to answer the question, and to create writing quality; it also reflects attention to the recipient's needs when giving feedback. It perhaps wouldn't be helpful to comment on handling of the question over and over again, to the neglect of other salient criteria. *Clarity* is the second most frequent code, reflecting the findings in the standards chapter of its growing importance over the course of the intervention.

Table 20: References to criteria codes in annotations and end comments on peer essays

	References in annotations	References in end comments	Total
Question (C21)	134	123	257
Clarity (C4)	98	106	204
Argument (C2)	65	112	177
Signposting… (C26)	58	56	114
Critics and theory (C9)	35	63	98
Introduction (C15)	27	62	89
Conclusion (C6)	21	50	71
Development (C11)	32	39	71
Analysis (C1)	31	38	69
Evidence (C12)	18	41	59
Style - concise and precise (C29)	31	20	51
Quotes (C22)	20	27	47
Knowledge, understanding, ideas (C16)	13	32	45
Consistency, coherence, focus (C7)	9	33	42
Range; scope (C23)	7	32	39
Voice, positioning, (C33)	8	27	35
Technical accuracy (C31)	22	13	35
Relevance (C24)	23	12	35
Context and period (C8)	17	17	34
Depth and detail (C10)	12	21	33
Comparison and linking (C5)	14	18	32
Structure (C27)	5	24	29
Style (C28)	8	19	27
General; broad (C14)	9	15	24
Technical – referencing (C30)	14	5	19
Nuance; complicating (C17)	9	4	13
Choice of authors or texts (C3)	1	11	12
Terminology (C32)	2	8	10
Finishing (C13)	3	7	10
Plot (C19)	4	6	10
Original (C18)	0	7	7
Specific knowledge (C25)	0	6	6
Progresses; direction (C20)	0	5	5
Word count and essay length (C34)	1	0	1
Total	751	1059	1810

Overall, the top ten most frequent criteria codes are consistent with those reported for the whole dataset as reported in Table 9, Chapter 5, with one exception: *evidence* is included instead of *quotes*. This reflects the higher frequency of overall references in end comments as compared to annotations: *quotes* appears more frequently in annotations, which are more likely to be detail-oriented than end comments. Similarly, there are many more references to *argument*, a broad criterion, in end comments than in annotations, while other, more detail-oriented codes are more frequent in annotations. For example, *style – concise and precise, technical accuracy, relevance*, *technical – referencing* are all more relatively more frequent in the annotations data. The very lowest frequency criteria reflect the overall dataset. Of note are *original* and *nuance; complicating*. In Chapter 5, these were mapped onto two Faculty criteria which are rather ambitious for first year students: 'intellectual penetration and sophistication of conceptualisation' and 'originality of argument'. Thus it is unsurprising that that these codes are low frequency, although it is perhaps encouraging that these qualities are noticed at all. Linking with the analysis in Chapter 6, this implies that students are beginning to make challenging criteria their own. An alternative explanation is that the peer essays are likely to have been of higher quality than the example essays, which were chosen to show a range, and mostly sampled from collections, sat in January and April, so earlier in the academic year.

8.3 What types of comments do students make when giving feedback?

In addition to coding for criteria, the feedback data were coded for type of comment, using a guide of the same name. As explained in Section 4.11.5, the coding guide built on work by Patchan et al. (2016), using four codes as a starting point: *praise, problem, solution* and *localisation*. Additional codes emerged inductively from the data and were used to capture comments which did not fit into this framework, particularly for annotations as Patchan et al. (2016) focussed only on end

comments. These were: justification for solution, problem-solution (where the phrasing was too complex to parse one or the other, value judgement ('overall, a strong essay' – Sean), and dialogic ('is it an example of Orwell's writing?' – Claire). As these codes are low frequency, for the moment, the analysis focuses on the three main comment types: praise, problem, and solution. The full coding guide can be found in Appendix L.

The definitions of Patchan et al.'s codes were also adjusted to the particular data and disciplinary context where necessary, as shown in bold below. For example, as reviewers rarely pointed out what was 'wrong' in writing, the definition of the code *problem* was expanded to encompass:

> Description of something that the reviewer **explicitly** considers to be wrong, **or detracting from the quality of the essay.**

In English, interpretations or ideas are rarely 'wrong', and a check on interview data showed that students rarely talked about right or wrong, or self-corrected when they did. Participants often highlighted problems more obliquely: 'In some paragraphs like paragraph 1, you give quite a lot of narrative about what happens in the text before you analyse why it is important' (Gemma). *Solutions* provided explicit suggestions for how to fix a problem or improve the quality of the essay, as for Patchan et al. (2016). However, as writers were not revising these essays, solutions were sometimes phased using past conditionals: 'could have used more historical context and maybe one more critical viewpoint in order to strengthen the argument.' (Chloe). Here, it can be easily deduced despite the 'could have' that the solution is to use more context and one more critic. *Praise* remained as for Patchan et al. (2016) – a positive feature of the essay ('I thought that the textual detail was great and well engaged with' – Abbie). A *localisation* is a comment that describes the location of a problem or, less often, and in this study only, a positive feature.

Each annotation and end comment was divided into segments, each consisting of a distinct comment. These segments were predominantly coded as *praise, problem*, and *solution*, and Table 21 below reports the overall coding frequencies. Unlike the criteria coding guide, only one code could be selected for each segment: codes were mutually exclusive. The exception was the code *localisation*, which was used in conjunction with the other codes if, for example, the position of a *problem* in the essay was given in the end comment. Further codes were used to capture other types of comment but these were low frequency. In order to maintain focus, they are not discussed here but can be found in Appendix L.

Table 21 shows that the split between *praise* and areas for improvement – *problems* and *solutions* – was fairly even, but with end comments providing more praise. This balance firstly reflects instruction in the workshops to take a developmental rather than deficit perspective, and the ground rule to provide constructive feedback. The higher proportion of praise in end comments reflects that annotations were usually made for the reviewer only. While problems and solutions were fairly evenly balanced in annotations, the emphasis in end comments shifts towards providing solutions. This suggests that reviewers were attempting to take critique and make it helpful to the writer, as instructed.

Table 21: Types of comment in annotations and end comments on peer essays

	Annotations		End comments	
	Freq.	%	Freq.	%
Praise	257	47%	345	53%
Problem	151	27%	123	19%
Solution	144	26%	177	27%
	552	100%	645	100%

Next, the analysis presented examines the distribution of praise-problem-solution comments across the criteria. As the focus of the present chapter and research question is on the role of received feedback in self-regulating writing, rather than noticing through giving feedback, the analysis will focus exclusively on the end comments. Participants could share their annotations with recipients if they wished, but few chose to.

Table 22 cross-tabulates praise-problem-solution against the criteria, and as before the criteria codes are ordered by total number of references. As each individual praise, problem or solution segment could be coded as multiple criteria, the total number of references does not match that in Table 20 above. There are around double the number of references, which suggests that two criteria could be identified in the average comment; NVivo does not allow this to be calculated. The split of references across praise-problem-solution remains consistent within a percentage point with those reported in Table 21.

Question, clarity and *argument* are the most frequent codes, and again, *question* is not outstandingly dominant, reflecting the inter-connectedness of criteria. Comments on *question* and *argument* reflect the overall pattern, with an even split between praise on one hand, and seeing these criteria as an area for development on the other. In contrast, there is less praise for *clarity*, and references are distributed almost equally between praise, problem and solution. *Introductions* attracted more praise than conclusions, which were the object of predominantly problem and solution comments. This pattern perhaps reflects the time constraints of the task.

Table 22: Cross-tabulation showing coding results for type of feedback comment by criterion

	Praise		Problem		Solution		Total
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	
Question (C21)	77	50.0%	31	20.1%	46	29.9%	154
Clarity (C4)	55	36.7%	49	32.7%	46	30.7%	150
Argument (C2)	77	52.0%	27	18.2%	44	29.7%	148
Introduction (C15)	47	55.3%	18	21.2%	20	23.5%	85
Critics and theory (C9)	52	63.4%	7	8.5%	23	28.0%	82
Conclusion (C6)	34	44.2%	22	28.6%	21	27.3%	77
Signposting… (C26)	27	37.5%	10	13.9%	35	48.6%	72
Evidence (C12)	32	61.5%	6	11.5%	14	26.9%	52
Analysis (C1)	34	66.7%	7	13.7%	10	19.6%	51
Development (C11)	10	22.7%	9	20.5%	25	56.8%	44
Quotes (C22)	22	55.0%	7	17.5%	11	27.5%	40
Range; scope (C23)	22	56.4%	7	17.9%	10	25.6%	39
Knowledge… (C16)	30	81.1%	5	13.5%	2	5.4%	37
Voice… (C33)	28	77.8%	5	13.9%	3	8.3%	36
Consistency… (C7)	20	57.1%	9	25.7%	6	17.1%	35
Depth and detail (C10)	16	61.5%	3	11.5%	7	26.9%	26
Structure (C27)	11	44.0%	3	12.0%	11	44.0%	25
Style (C28)	13	59.1%	3	13.6%	6	27.3%	22
Comparison, linking (C5)	12	60.0%	5	25.0%	3	15.0%	20
Style – concise, precise (C29)	18	90.0%	2	10.0%	0	0.0%	20
General; broad (C14)	7	38.9%	10	55.6%	1	5.6%	18
Context and period (C8)	12	70.6%	0	0.0%	5	29.4%	17
Relevance (C24)	5	33.3%	7	46.7%	3	20.0%	15
Technical accuracy (C31)	3	20.0%	5	33.3%	7	46.7%	15
Choice of authors, texts (C3)	9	75.0%	1	8.3%	2	16.7%	12
Plot (C19)	4	40.0%	5	50.0%	1	10.0%	10
Terminology (C32)	5	50.0%	1	10.0%	4	40.0%	10
Specific knowledge (C25)	4	50.0%	3	37.5%	1	12.5%	8
Original (C18)	7	100.0%	0	0.0%	0	0.0%	7
Technical – referencing (C30)	3	42.9%	2	28.6%	2	28.6%	7
Finishing (C13)	2	33.3%	3	50.0%	1	16.7%	6
Nuance; complicating (C17)	4	66.7%	1	16.7%	1	16.7%	6
Progresses; direction (C20)	2	66.7%	1	33.3%	0	0.0%	3
Word count, length (C34)	0	0.0%	0	0.0%	0	0.0%	0
Total	704		274		371		1349

Development and *signposting* have the highest proportions of solution comments, followed by *technical accuracy* and *structure*. The criteria mapping in Chapter 5 showed that *signposting, development,* and *structure* are all linked to *clarity, question* and *argument,* whereas *technical accuracy* contributes only to *clarity*. These codes reflect procedural aspects of writing which are nonetheless important for key aspects of quality. It is perhaps more realistic for student writers to both offer and enact solutions for these criteria.

Many of the codes which predominantly attract praise are related to knowledge and material, for example: *knowledge, understanding and ideas, voice, positioning and opinion, analysis, critics and theory* and *evidence.* This may be for several reasons. Firstly, the emphasis of the intervention was on how texts are constructed, rather than subject content; students were steered away from critiquing content in discussion, but in peer feedback could give a personal response, and as the interviews show, students held great interest in their subject. Secondly, participants reviewed essays written by participants from other colleges. Shared knowledge cannot be assumed, as for these period papers, college tutors decide which authors to teach, and students select which Faculty lectures to attend. Therefore, it is likely that students would not have read the same texts, and therefore reviewers would be less likely to point out problems or solutions with knowledge. Such a novice perspective is also reflected in the coding of *original,* the only code to be derived from Faculty criteria. *Original* is low frequency, with only seven references, and coded exclusively as praise, suggesting that originality is a desirable trait but one which is infrequently identified in first-year writing by first-year students. Indeed, the Faculty criteria are intended for use throughout the degree course; student criteria present a more realistic description of first-year writing. *Nuance; complicating* is even lower frequency, with only six references, but is not solely praised, which suggests that that nuancing or complicating points and ideas is a more realistic for first year students, but again, is not noticed frequently in writing. Conversely, *general; broad*,

almost an opposite to *nuance* has the highest proportion of references coded as problem. Taken together, these results suggest that in some instances, participants identified knowledge-related strengths, ideals and pitfalls in writing, but the coding does not show that they had knowledge-focussed solutions to bridge them.

8.4 How do participants respond to feedback?

The chapter now turns to how participants responded to their feedback. Participants' initial responses are captured in the annotations they made on feedback, as well as reflections, which are discussed below. Twelve participants made annotations. These were coded using a further guide, 'getting feedback', and the codes emerged inductively from annotations. The results are summarised in Table 23, and are ordered by frequency of annotations. The coding here is not mutually exclusive. For example, 11 of the references to *action*, the second most frequent code, are also coded as *agree.*

Participants noted their agreement more frequently than their disagreement. Agreement was often voiced succinctly for example as 'Yes!' (Celine), 'fair' (Megan), or even ticks. Disagreements tended to be more extensive, and might question the reviewer: 'I thought that everything I did stemmed back to the quotation?' (Phoenix). The annotations also highlighted students' strategies for using feedback: *action* is the most frequent code, and six participants also used the strategy of *summarising feedback*, drawing out possible actions from multiple reviews. Students also noted where there were *conflicting reviewer views* or *reviewer consensus*. There was only one instance of a participant noting that feedback provided *new information*, and that related to *praise* received by Phoenix for a paraphrasing strategy. Participants also voiced affective responses to feedback, such as Rachael's 'Aw what an encouraging comment!' The writer's privileged,

metacognitive view on the process behind the product is apparent in *self-explaining*, for instance, Celine's 'fair point (I ran out of time so knew this was pretty weak)'.

Table 23: Types of annotations made on feedback received

Type of annotation on feedback	Participants	Annotations coded
Agree with reviewer	12	59
Action	12	47
Summarising feedback	6	28
Self-explaining e.g. process	12	23
Disagree with reviewer	10	19
Affective	5	10
Conflicting reviewer views	5	5
Reviewer consensus	5	5
Dialogic response	3	4
New information	1	1

Tracing annotations back to types of comment reveals no particularly strong patterns for criteria: participants tended to annotate the highest frequency criteria, as expected. However, linking annotations to type of comment does provide some insights. Table 24 cross-tabulates the main 'getting feedback' codes with the type of feedback to which students were responding, i.e. 'praise-problem-solution'. Although *praise* comments account for around half of the feedback comments, they were less frequently annotated than *problem* and *solution* comments. *Solutions* were the most frequently annotated, and also the most frequently disagreed with. For instance, Gemma questioned whether 'linking back to the question and your overarching response to it at the end of paragraphs' could be 'a bit excessive? – every paragraph feels unnecessary'. In 11 instances, students planned *actions* for *problems*. For instance, in response to a comment that 'sometimes the links between texts and paragraphs were tenuous', Megan responded, 'I needed to spend more time planning to complete tease out how to link each text.' This suggests that participants

were able to propose their own solutions; nonetheless, as the next section shows, analyses show that solutions-focussed feedback was considered more helpful.

Table 24: Responses to praise, problem and solution comments

	Praise	Problem	Solution	Total
Agree with reviewer	5	17	30	52
Disagree with reviewer	0	6	8	14
Action	2	11	22	35
Total	7	34	60	

8.5 What defines helpful feedback?

It can be assumed that in order to use feedback, students need to be receptive, and find it helpful. This section discusses the nature of helpful feedback in this study. Post-feedback reflections gathered students' perceptions of the comments received, with one item asking 'Which reviewer gave the most helpful comments and why?' All participants responded to this item, and the data were analysed in two ways. First, to identify whether the comments cited as most helpful shared any characteristics, the end comments were identified and compared as a group against all end comment feedback (n = 66). Two participants each cited two reviewers as the most helpful, and so 23 sets of end comments were included in the analysis. Secondly, participants' responses were analysed qualitatively, using a separate coding guide, helpful/unhelpful, found in Appendix L. The mean word count of responses to the reflections item was 45 words (SD = 19.9).

Interestingly, reviewer helpfulness was distributed across the participant group: 17 of the 21 feedback-givers had at least one piece of feedback rated as most helpful. Only one reviewer, Poppy, had all three pieces rated as most helpful. Four others were cited for two pieces: Jess, Rachael, Sean and Tessa.

Table 25 breaks down the proportion of comment types into three columns: all end comments, most helpful end comments only, and those not rated as most helpful. *Praise* is the most frequent type of comment across the dataset, but the most helpful subset contained a lower proportion of *praise* comments, and also *problem* comments. Instead, the most helpful feedback had a higher proportion of *solutions,* and the row is highlighted to point this out. Other comment types which are low frequency but which have salience in the literature are also included. *General overall comments*, which include judgemental language but don't draw on a particular criterion are included, such as 'overall, a strong essay' (Sean). Such comments are not considered helpful in the literature as they communicate a sense of finality (Boud, 1995). Dialogue, too, is increasingly important in the feedback literature (e.g. Nicol, 2010), and a small number of comments coded as *dialogic* tried to start a conversation with the writer, often through questions: e.g. 'Is it an example of Orwell's writing?' (Claire). However, neither of these codes seemed to contribute (or not) to perceived helpfulness.

Table 25: Types of comment across subsets of feedback data

	All end comments (n = 66)		Most helpful end comments *only* (n = 23)		End comments *not* rated as most helpful (n = 43)	
	Freq.	%	Freq.	%	Freq.	%
Praise	345	48.4%	138	45.5%	207	50.5%
Problem	123	17.3%	47	15.5%	76	18.5%
Solution	177	24.8%	87	28.7%	90	22.0%
of which localisations	*88*		*39*		*49*	
Dialogic	29	4.1%	14	4.6%	15	3.7%
Justification for solution	14	2.0%	7	2.3%	7	1.7%
General overall comment	25	3.5%	10	3.3%	15	3.7%
Total	713		303		415	

Reflections responses align with the analysis in Table 25 to confirm that the most helpful feedback was more focussed on solutions. In the reflections data, the most helpful reviewers were most frequently cited as offering suggestions or explaining how the writer could improve (9 cases). Participants also appreciated comments that were specific rather than general (9 cases). Indeed, for Chloe and Megan, the two case study participants who directly acted on feedback, the peer suggestions they implemented were both solution-focussed and specific. For example, Megan was 'already aware' of the need for clearer 'lead in' or topic sentences. For her, 'the feedback was quite good as it kind of like [pause] showed me what else I could do'. Holly, the reviewer who made these suggestions, and whom Megan rated as most helpful, provided annotations, and also identified the 'theme [that could] could be leading into a sentence'. This is a different topic sentence strategy to the one Megan was using; as she notes on her feedback: 'The feedback person is really not a fan of aesthetic lead in'. However, despite these differences, Megan sees the suggested topic sentence approach as a solution to an issue pointed out by another reviewer:

> 'Reviewer 1 does not think I am addressing the question well enough – perhaps I need to use topic sentences more as reviewer 2 *[i.e. Holly]* suggests to make things more explicit.'

Megan partially agrees with Reviewer 1's critique. Although in when writing she had 'tried to explicitly reference the quotation', in re-reading her essay after giving feedback, she noticed that 'I don't think I am explicit enough with the question' (Workshop 2&3 Reflections). This again reflects the conflict that students sometimes experience between the quotation and the question stem, as discussed in Chapter 6. For Megan, using more 'focussed' topic sentences in her Interview 2 essay is also in the service of another goal: time management. She also changes the structure of her essay, using shorter paragraphs to ensure she finishes. Adapting her topic sentences, then, allows her to fulfil multiple purposes.

Although the most helpful feedback contained proportionally fewer praise comments than other feedback, praise was appreciated by participants throughout the datasets to be helpful. For example, Megan annotated a praising end comment with ‘ahh cheers’. In reflections, six writers noted appreciation for the positive tone or perceived attitude of the reviewer; as Amber put it, her favoured reviewer was ‘striving to be helpful’. In Interview 2, Holly commented on how a reviewer who had provided annotations and ‘specific’ feedback conveyed an appreciative stance:

> ‘Looking through it, it’s always it’s nice to have some positive feedback as well and even when it’s just going down the essay and just seeing ‘good’, ‘good’, ‘good’, you might know it’s good but just to hear that every couple of you know, yeah. It’s - it’s positive and it makes you take in the rest of the feedback a lot better when you know that someone’s actually reading though your work and appreciating what you’re doing.’

For Holly, praise confirmed her overall judgement that the essay was ‘good’, but also helps the reviewer-writer relationship, and her receptivity to feedback. For other case study participants, praise comments provided more specific ‘verification’ about which aspects of writing worked, and should be continued.

> ‘But it was quite nice as well, where it sort of showed you things that you did well, because then you could sort of see 'OK, this is obviously working, I'll try and keep that' (Tessa, Interview 2)

> ‘It was nice as well, the reviews, because although it gave me points to work on, overall it was quite... quite complimentary of the essay, which meant that I'd had kind of verification that was a good essay to do. […] and if I know what a good essay is, I can take that into something else.’ (Phoenix, Interview 2)

It is easy to assume that feedback should always provide next steps, but taking stock and deciding what doesn’t need to be adapted is important for self-regulating learners, too. This need for a ‘balance of positive comments and constructive criticism’, as Zoe put it in reflections, was also highlighted by four participants from the wider group. In fact, two writers found some reviews

overly complimentary. This is discussed in the next section, which deals with the features of feedback considered to be unhelpful.

8.7 What makes feedback unhelpful?

Alongside the reflections item on helpfulness, participants were also asked, 'Were any of the comments unhelpful? If so, please explain'. The mean word count for this item was slightly lower than the word count for the 'helpful' item, at 37 words (SD = 17.8) as opposed to 44 (SD = 20). 13 of the 21 students gave examples in direct response to the question, and 2 students gave examples of comments they considered 'perhaps less relevant' (Heather) or 'not especially' unhelpful (Rachael). The reasons cited tended to be in direct opposition to what participants considered 'helpful'. Students pointed to instances where feedback was insufficiently specific, or didn't include an example or way to improve (9 cases). In three cases, a reviewer's ideas about content or argument differed to the recipient's, and the reviewer did not take this into account. As Poppy states, commenting on argument '[it] is a matter of interpretation anyway, so there could be a more constructive way to comment on the essay'. There were only three instances coded where participants explicitly pointed out the limitations of the reviewer's perspective as a student, for example: 'Some terms 'expand on this' a bit vague, but appreciate that student markers may not know <u>how</u> to expand it' (Jess, her emphasis).

As English students skilled in close reading, participants paid close attention the wording of comments in their reflections. A key code was 'how the reviewer comes across', which pertains to tone, and/or where a reviewer does not come across as trustworthy. Examples are:

> '"This was a really interesting essay and choice of question" was a bit generic and bordering on sympathy' (Sean)

> 'Comments like "very strong" and "quite intelligent" seemed a little biased and potentially less helpful. "Sake of the reader" also seemed a little condescending to me, though I appreciate the thought behind it' (Abbie)

> 'Though I agree with much of it, it's quite blunt and doesn't necessarily point out ideas for improvement as much as the other 2' (Rachael)

In all three examples, the reviewer's tone and stance disrupted the symmetrical, helpful peer-to-peer relationship. In Sean and Abbie's cases, the reviewers disrupted the symmetry between reviewer and writer by offering 'sympathy' or 'condescens[ion]'. As a result, praise was not taken as praise as it did not appear genuine, but rather a 'generic' or filler comment. As Molloy et al. (2013) observe, such comments may be intended to soften criticism, but do not in fact offer the information on performance that recipients really want and expect from feedback. Although such comments might be intended to pre-empt and mitigate against undesirable emotions, they can actually have the opposite effect.

In Rachael's case, the peer-to-peer relationship was undermined by 'blunt' comments. Likewise, Ben received 'blunt' feedback which he described in Interview 2 as 'awful' and 'cutting'. This is given in Appendix G. Although Ben agreed with some of the comments, which were corroborated by other reviewers, he was 'affected' by the tone:

> 'It's just really like blunt. I – it's really odd that it was the tone that affected me more than the comments, because the comments were right, it's just the… the way something is articulated makes such difference. Umm [pause, reading]. Yeah, so like 'several points were not properly explained and therefore risk subverting your argument'. Like, 'pay attention to syntax', they were all quite like concise, or very kind of yeah, blunt comments. […] 'You do not give sufficient justification for the position you take.' I was like, OK, sorry! Whereas the other people were saying like, very complimentary stuff and often it was conflicting with what reader 4 thought. Or sometimes it was the same, but it was said in a different way, but you read the other ones and they're a lot easier to deal with'

Whereas in the previous section reviewers were perceived as appreciative readers, here the peer-to-peer symmetry is again disrupted, with the 'blunt' comments forcing Ben into a position of apology: 'I was like, OK, sorry!' Solutions are not given as suggestions but rather as imperatives – 'pay attention to syntax', mirroring findings in the wider feedback literature that a harsher, more directive tone is used for weaker work (Wingate, 2010; Court, 2014). And yet this work is not graded – it cannot be explicitly positioned as weak – and the reviewer is not a teacher. However, in their tone and framing of the work as problematic, the reviewer is constructed as an authority, passing judgement.

As well giving rise to unpleasant emotions, such feedback that invokes final, absolute judgements risks undermining the agency and self-esteem of the recipient (Boud, 1995). However, as Ben had received three other pieces of feedback, he was able to contextualise the 'blunt' feedback as an outlier, or the product of an individual, subjective reading: others were more 'complimentary' or at least 'a lot easier to deal with'. He also explains how over his first year at Oxford, he had 'got used to certain like… methods of criticism which are quite cutting and quite blunt' – what he terms the 'house style' of particular college tutors:

> 'Although… it's cutting on a personal level, it feels like it's cutting on a personal level, you've got to realise that it's about the text, it's about the essay. And like you've got to take the points out of that in the same way that you've got to take, you've got to remove yourself from all the feeling of how good you are from other people's responses and find the bad things and work out how to improve this'

None of the feedback in this study would directly fall into a research definition of person level feedback, which would involve a comment such as 'you are a good writer' (Hattie and Timperley, 2007). However, Ben's experiences reflect observations in the literature that it can be interpreted as such by recipients (Boud, 1995; Molloy et al., 2013). Ben has learned to extract information from feedback that 'feels like it's cutting on a personal level', reappraising it to be 'about the

essay'. It is still up to him, though, to 'work out how to improve this', as such feedback does not suggest helpful ways forward. This feedback, then, as well as placing an affective burden on the recipient, also does not offer guidance on where to go next. So, while the feedback-giver takes an authoritative stance in terms of tone and judgement, they do not actually take on a role where they could co-regulate the recipient's learning, or provide useful input to self-regulate learning – what to adapt and do differently next time.

As described in Section 4.10.3, two further comments had been redacted from Ben's feedback as I held ethical concerns about the potential impact of the harsh tone: as Torrance (2012) points out, assessment can be 'deformative'. Ben also received four pieces of feedback - to mitigate against the risk of one student being unable to provide feedback due to illness, three students reviewed one extra essay: Ben's was one of these. While the redacted comments would have been coded as solutions, the suggestions were not helpful and condescending. For example, the italics here were redacted: 'argument appears confusing in the introduction – *be direct and convincing by arguing something that makes sense'*. This seems to be a problem – the argument doesn't make sense – dressed up as a solution. Ben's problematic feedback was written by Heather, who in her Workshop 4 reflections responded that she was surprised by the 'level of positivity' in the feedback she received. Despite the instructions to take a developmental perspective on writing, perhaps Heather saw robust criticism as a way of helping the recipient. Indeed, Ben's comments on 'house styles' of critique suggests that the tone of her comments could reflect a wider disciplinary and university culture of critique. Two participants commented in reflections that the feedback from one or two reviewers was too complimentary: as Gemma comments, 'their feedback <u>was</u> useful at points, I just wish they were a bit <u>more</u> critical' (original emphasis). Joe is more direct: 'I actually think that reviewer 1 & 3 are hampered by how polite they are'. In contrast, his most helpful reviewer 'they didn't spend too long complimenting and got straight to

the useful and constructive feedback'. He is almost suspicious: 'I also feel like some of the compliments carry implied criticism, so perhaps a bit more clarity and honesty?' In the Oxford environment, a direct, less 'polite' critique from tutors may be possible due to the working relationship, developed over a term or even academic years. In interviews, students referred to conversations with tutors which may not be so frequent or widely available in other systems; or students may not feel able to seek such conversations. In contrast to these usual modes of feedback, peer assessment was based on anonymous, written feedback, which neither allowed for the dialogue nor for trust between reviewer and recipient to be built up, both of which are vital, particularly where feedback is challenging to the learner (Carless, 2013a; Nicol, 2010). Without such dialogue, robust critique is likely to lead to dissatisfaction or even distress for the recipient. As Ben's experience shows, such feedback is likely to be helpful to learning only if the recipient has strategies to get past any comments on the 'personal level', interpret the message of the feedback, and has an idea of the next steps they can take to improve. This is highly demanding. In other cases, writers may not have such knowledge and strategies.

8.8 Agreeing with but not implementing feedback

In order to examine why writers might chose not to act on feedback with which they agreed, and which was corroborated by multiple peers, this section begins with Phoenix. As Figure 19 below shows, all three reviewers commented on the need to develop the conclusion, and Phoenix's underlining and annotations – in blue – show his agreement. Phoenix was already somewhat aware of this issue, commenting immediately post-writing: 'I felt it was a little rushed at the end'. His post-feedback reflections highlight what he perceived to be the key message of his feedback: [the] 'need for a strong conclusion – I didn't realise <u>how much</u> of an issue a quick conclusion was.'

Figure 19: Feedback to Phoenix on his conclusion

Reviewer 1 (Abbie)	Saying this the argument was still developing later in the essay and the ultimate conclusion felt like a powerful one which perhaps just deserved some more space to unpack it?	**Make sure I reserve time for a conc.**
Reviewer 2 (Poppy)	Introduction showed careful consideration of the terms of the question, and established a sense of intent; a good foregrounding for the essay. The conclusion could have been similarly potent – a nice example to finish but perhaps a more concise, declarative final assessment would have invoked more finality.	**Make sure I always do this.** **Emphasise need to focus on a strong conc. → save time**
Reviewer 3 (Tessa)	– Your conclusion seems a bit abrupt. Interesting point about subjectivity – explore its implications.	**Again yes.**

When recalling his feedback in Interview 2, it is the first criterion he mentions: 'they were like completely true about the need for a conclusion, and I think that is a big thing'. However, he doesn't manage to write a conclusion at Interview 2, and also notes, tellingly, that 'I don't think I've ever written an essay where I've been able to write a nice conclusion at the end'.

Phoenix explains that time – and the research situation – means he does not manage a conclusion in Interview 2. However, he is not struggling for words – his essay is amongst the lengthiest in the dataset. He does think he will manage a conclusion in the exam itself. As the extract below shows, Phoenix's goals are likely to be different in the 'actual exam'. The Interview 2 essay is exam preparation, working through material, and so his standard for finishing is idea-led – '[getting] everything I wanted in'.

> 'I finished the essay in the sense that I got everything I wanted in, I didn't write a conclusion because of time limits, but I reckon in the actual exam both

> adrenaline and the fact I won't be speaking out loud in my plan means I'll have that extra three minutes, which is all I needed to do a summative conclusion, because I always put the bulk of my essay summary in the intro just in case that happens.'

The three-minute 'summative conclusion' here seems to contrast with his reviewers' perceptions of introductions: Poppy's 'declarative final assessment' or Tessa's 'explor[ing]… implications'. The latter term is shared by other participants from Tessa's college, too, who were taught this move by a college tutor. For Phoenix, however, the specific rhetorical contribution of the conclusion is less clear. In Interview 1, he explains that 'I like my introductions to like, be like my conclusion, to encapsulate my argument, so that if I don't get to a conclusion because of time… they know what my argument is and what my contention is.' While Tessa's 'implications' suggests a progression, as does Abbie's point about 'the argument… still developing', for Phoenix, the introduction has already accomplished the purpose of communicating his 'contention'. Thus, while the reviewers convince Phoenix of the need for an conclusion, his conception of what a conclusion is and does, in contrast, does not appear to shift over the course of the study.

However, Phoenix does make a related adaptation in the Interview 2 essay: he 'narrows… down' his material, leaving out 'some of my lovely points'. This adaptation seems to be a forerunner or prerequisite to making space and time for a conclusion. In Interview 2, hypothesising what he would do with an extra ten minutes, Phoenix voices conflicting priorities. While he first states that he would have 'written a conclusion' and edited the introduction, he also acknowledges that:

> 'But saying that, I know me and I also want to try and fit in as much as I can, so I might be stupid and try and put in another point. But I hope by now I realise that that's not a good idea.'

Phoenix's comment highlights how writers experience conflicting standards, and particularly the strong pull of those rooted in enjoyment and interest in the subject. Such standards may be long-

held, and rooted in old writing habits, for example Phoenix's use of his own quotation to open the Interview 1 essay, as described in Section 6.3. Here, Phoenix is torn between writing about as many of 'my lovely points' as possible and writing a conclusion to meet his readers' expectations. With metacognitive knowledge of his writerly tendencies – 'I know me' – he can choose whether follow his own inclinations, or conform to the expectations of his readers. 'Narrowing down' material is a step towards the latter, and perhaps a more achievable adjustment towards writing a conclusion.

Phoenix, then, although amenable to feedback, retains strong ownership of his writing, as participants do throughout the dataset. In Holly's case, her sense of ownership was more strongly voiced. Holly thought her feedback was 'insightful' and was impressed with the way her reviewers 'analysed it in depth' (Interview 2). As with Phoenix, all three reviewers commented on one area: the need for further signposting in the introduction. In reflections, she reflects that 'I think this is definitely worth noticing because it is something an examiner might pick out that I would not otherwise have discerned.' She lists it as feedback she would consider acting on. However, in the run-up to Interview 2, she reflects on whether taking action is worthwhile, deciding not. She had already adapted her writing to include one signposting move – she tries 'this essay will' for the first time in the Interview 1 essay – and continues with this adaptation. However, she isn't comfortable with adding another to 'lay out' her texts:

> '…that's come back to how do I approach my introduction, do I lay out my texts and where I'm going or do I work it in as I go along? I think that's something I've been adjusting to, while coming to uni, because before I didn't used to write 'this essay will explore', 'this essay will argue that', that's something that I've been taught now, and it's almost I can put that in so I know where I'm going, as, I suppose, sort of a pointer. It does still feel a little bit awkward, but I've almost got to the point where I've forgotten what I was doing before, but I don't think I can sort of take the 'and I'm going to write about these authors and I'm going to write about these authors because...' and the dates, I think that's still further away from my writing style than I'd like it to be.'

It is interesting that the move Holly has adopted – 'this essay will explore' – is not solely for the reader: it works as a 'pointer' to clarify the argument for her as a writer, too. However, explicitly using this move remains 'a little bit awkward' as it has not quite yet overwritten her previous way of writing introductions. She still much prefers to 'work… in' the texts as she goes along: this is her style. Later in the interview, she explains that making too many changes to writing style is risky for the integrity of the essay:

> 'I think also if you stray too far away from your writing style, it can kind of compromise the rest of your essay. It's better to try and stick with yourself, and it's not necessarily hundred percent what other people want, but then it's never going to be.'

For Holly, then, there is a limit to how much she will adapt her writing style to please readers. As with Phoenix's conflict between points and conclusion, this links back to the tension that Nicole voices between creativity and criteria in Section 6.12. Holly acknowledges that her peers' views on signposting in the introduction highlight an issue than an examiner could also find problematic. However, her commitment to her own style is more compelling. She resists the convergent message of the feedback – including a conventional move in the introduction – recognising that judgement varies between readers.

8.9 Feedback as a 'boost' to self-efficacy

A final, and important, impact of feedback for some writers was on self-efficacy. As described above, participants voiced their surprise at the positivity of the feedback, and praise helped writers decide which aspects of their writing they should continue with. For two participants in particular, peer feedback was a source of self-efficacy. While Holly and Phoenix above found that feedback confirmed that self-evaluation that their essays were 'good', others' self-evaluations

were less positive. As Nicole put it in her Workshop 4 reflections, 'it is a boost to have largely positive feedback, as I focused mostly on the negatives in my own review.' Similarly, Felicia's feedback offered new information, and so was 'quite encouraging':

> 'It was also quite encouraging that actually I can do this […] I don't know – often when you, you're doing the work on your own, you don't really know where you are, you can't really gauge, so yeah, it was good to be able to have feedback and get more of an idea of like the level that, if I was writing a decent essay or not really [laughs]'

The feedback Felicia receives allows her to more judge with more confidence 'where you are'. The essay isn't quite a mastery experience, but feedback does show her that her essay is 'decent' and 'actually I can do this'. This last phrase is echoed by Chloe, whose feedback was 'really nice', and highlighted 'strengths' in her work:

> Chloe: The strengths of my writing, I wasn't aware that there were strengths [laughter] so that was nice. Umm but it was probably just on this particular essay, but yeah. So that was [trails off]
>
> Researcher: So it was nice to have someone sort of point out what's good about it?
>
> Chloe: Yeah, it was like 'you can do this', sort of like 'oh really?' [laughs] But yeah.

Chloe was not alone in having difficulty articulating the strengths of her writing: across the group, it was difficult to elicit answers from the otherwise fluent and eloquent participants to the related interview question. Chloe doesn't appear to be entirely persuaded by her reviewers - it is interesting that she immediately downplays the presence of strengths – 'but it was probably just on this particular essay'. Her 'oh really?' also conveys her surprise at the reviewers' message that 'you can do this'. However, Chloe is aware of her self-criticism: elsewhere in the interview, the first point she makes when describing her workshop learning is that it has raised her metacognitive

awareness that 'I'm really critical, both of my own essays and other people's'. In her initial description of her feedback, she also cites one possible source of her critical self-evaluations:

> I thought it [the feedback] was really nice, I was really surprised! [..] I thought it [the essay] was OK, but I just didn't think it was… I'm clearly not used to getting good feedback [laughter] but umm... err, yeah, I was quite – I was quite surprised, and I also thought it was so helpful and really relevant.

Tutor feedback was beyond the scope of the study, but Chloe's comment suggests that the feedback she usually receives is critical: 'I'm clearly not used to getting good feedback.' Even if feedback is not intended in this way, it is often perceived to be so, as in the case of Ben's feedback, above. However, unlike some participants, Chloe doesn't discount the peer feedback for being 'nice' – it is also 'so helpful and really relevant'. As described above, it gives her ideas for tackling 'areas that I needed to work on', and she tries out some of the strategies recommended.

For students with lower self-efficacy beliefs, then, feedback may be a valuable source of self-efficacy and make the challenging, consequence-laden task of exam writing seem less daunting. However, it is important to note that in Chloe's case at least, peer feedback was combined with the other elements of the intervention: she also emphasises the benefit of looking at others' essays. This will be discussed further in the next chapter.

8.10 Chapter conclusion

This chapter has shown how self-regulating writers interpreted and used peer feedback. Writers in this study valued solutions-focussed feedback. Overall, student perceptions of feedback were positive, and writers' annotations on feedback voiced their agreement more often than disagreement. Nonetheless, only two case study writers directly implemented peer suggestions. In some senses, given the quality and volume of the feedback, this is surprising. Yet feedback was not

the sole source of information available to students to inform judgements and decisions about what to do next.. Through assessing and discussing example and peer essays, and reflecting on their own writing, students had already added to their metacognitive awareness of writing. Low uptake of peer feedback is a common phenomenon in the literature, even when students are given the opportunity to use peer feedback to revise work which is later graded (van den Berg et al., 2006; Walker, 2014; Patchan et al., 2016). This chapter has shed light both on why students don't enact peer feedback, and also of the other uses to which they put feedback.

Throughout this study, and in this chapter in particular, students voiced strong ownership of their writing, and in some cases this made it less likely to feedback to be implemented. As Holly highlighted, changes to writing can be 'awkward' and have to be enacted incrementally: 'stray[ing] too far away from your writing style' could be risky. It is also not always possible to immediately implement peer feedback, as Phoenix showed. Before making space for a conclusion, he needed to work towards making a narrower selection of 'my lovely points'. From a reader's perspective, including a conclusion seems straightforward, but writers may need to take interim actions in order to transition away from long-held habits. There may be several developmental steps to take before the feedback is feasible; such steps may be invisible to a reader-reviewer, but all too apparent to the writer. In addition, as we have seen in previous chapters, time posed significant barriers to adaptation and to feedback uptake. The implications will be discussed in the next chapter.

Feedback often confirmed students' own judgements of their writing. Does this high level of awareness make feedback redundant, then? The analysis in this chapter suggests not. Feedback also served provide 'new ideas about things that I was already aware of', as Heather put it. In particular, praise comments that highlighted genuine strengths of writing emerged as important.

For Chloe and Felicia, the 'new ideas' – the strengths of their writing which were so hard to see – provided a boost to self-efficacy, a sense that 'actually I can do this' (Felicia). For other participants, praise provided 'verification' of which aspects of writing to keep, to transfer to the next essay (Phoenix). Adaptation is not just about change – in a time-limited situation, information about which aspects of writing worked allowed students to prioritise and so self-regulate more productively.

Not all participants welcomed praise, and some comments offered 'sympathy' rather than information (Sean). Through the contested issue of praise, the disciplinary and institutional setting – itself not a direct object of the study – crept in. Some students were perhaps used to more robust critique from tutors. Applying close reading, Joe was suspicious of praise, which could obscure the real message: 'I also feel like some of the compliments carry implied criticism, so perhaps a bit more clarity and honesty?' Equally, the comments Ben received were perhaps intended to convey such 'honesty', but the 'cutting' and 'blunt' tone undermined the message. The potential for distress was mitigated by Ben's ability to go beyond 'the personal level', and the provision of other reviews, which enabled him to contextualise the comments.

This chapter has examined how students interpret and use peer feedback, addressing in particular why there is not often a linear progression from comment to enactment. These findings will be further contextualised in the next chapter, both in terms of the theory of action and the wider literatures.

Chapter 9
Discussion, Implications and Conclusions

9.1 Introduction

This study has examined the use of holistic peer assessment workshops to support writing in the discipline of English literature, and the findings are promising. Eight of the ten case study writers made adaptations to their writing which could be linked to their learning during peer assessment. Assessing example and peer essays gave rise to a wide range of criteria, which overlapped well with Faculty criteria, and additionally highlighted procedural criteria which students could use to develop their writing. Students' own self-set metacognitive standards shifted over time towards being more reader-oriented, and the way students defined and interpreted the task began to include the complex quotations that accompany questions. Feedback led to very few direct implementations of suggestions, analysis showed how students interpreted comments, and how feedback served other functions. The study has also shown that holistic peer assessment has a positive impact on ratings of self-efficacy, and has acknowledged the role of emotion in writing and assessment.

This chapter brings the findings together, recontextualising them with theory and literature to consider the implications and contributions of the study. Rather than dealing with the study's strengths, limitations and implications in separate sections, these are kept in context and dealt with as they arise. The chapter begins by revisiting and updating the theory of action with conditions which arose as important during analysis. Next, I examine what the findings imply about evaluative expertise, connecting these to students' own shifting standards and to the wider literature. The discussion then considers the role of peer feedback where students have developed evaluative expertise, considering how the interpretations and uses of feedback in this study fit

with theory. Next, I review the impact of the workshops on students' self-efficacy beliefs, which presents implications for the transition to university. I then revisit the practical problem of supporting writing within the discipline and whether the declarative evaluative expertise about quality is sufficient for developing writing. This leads into implications for practitioners. The chapter then rounds off the thesis by drawing together final conclusions.

9.2 Revisiting the theory of action

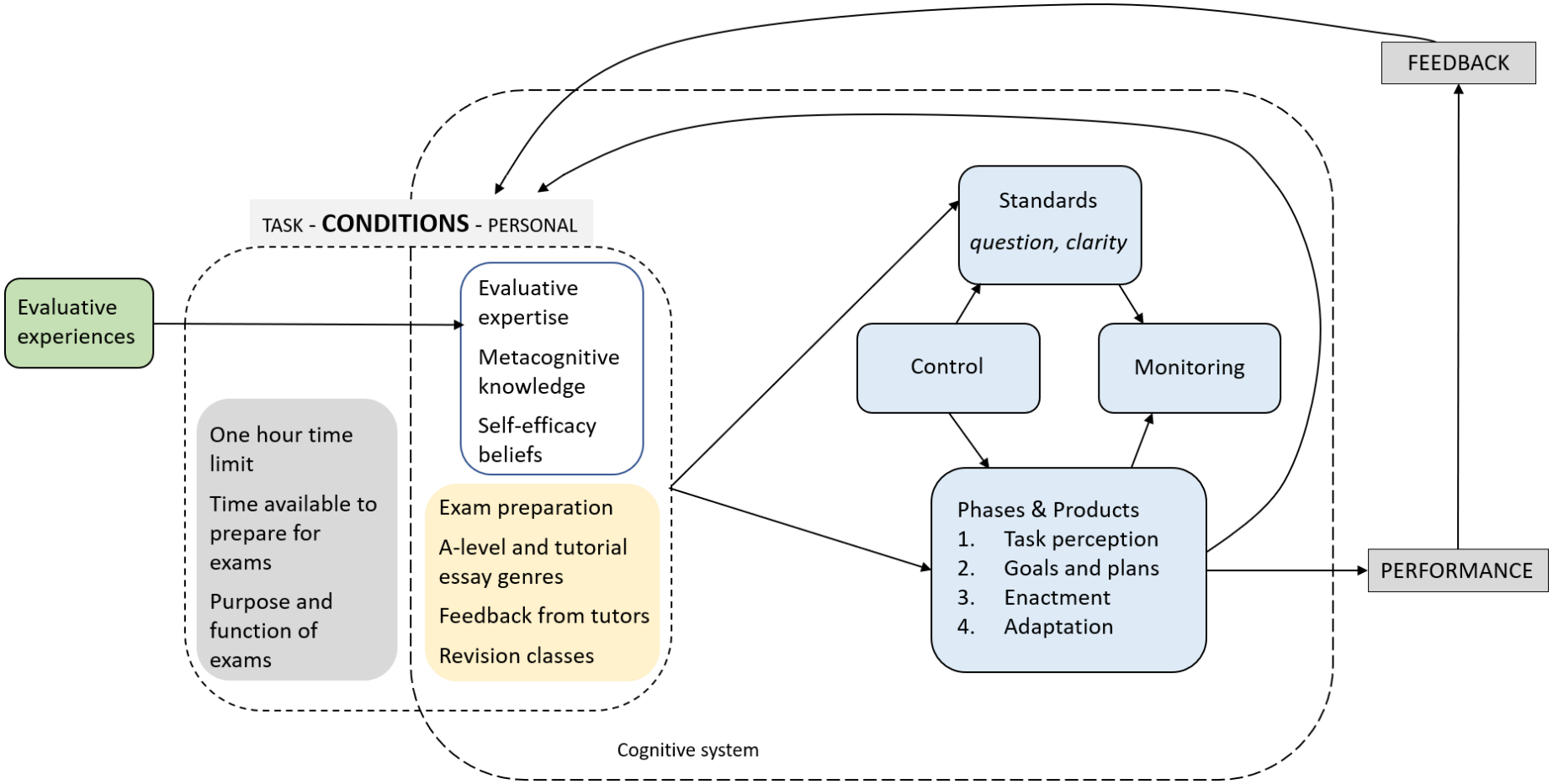
A key contribution of the study is to examine in detail the connection between assessment and self-regulated learning at the level of individual students, and over some time – four months. SRL is an area of growing interest in the assessment field (e.g. Nicol and MacFarlane-Dick, 2006; Black and Wiliam, 2009). However, representations of SRL in the higher education literature tend to frame the relationship as students *becoming* self-regulated learners: SRL is presented as an either/or, whereas in fact students self-regulate from a young age, given the opportunity (e.g. Perry, 1998). Many higher education papers treat SRL as an aptitude, whereas SRL theorists see it as a process (Winne and Perry, 2000; Zimmerman, 2008). In line with this shift, the present study found that students both self-regulated during the writing process and adapted their approach from essay to essay. It is worth, then, shifting the emphasis from *whether* students self-regulate to *how* students self-regulate, and whether this is productive and/or aligned with the learning valued in the context. Students may self-regulate in a ways that teachers might not value. In this study, this can be seen in students' over-emphasis on material rather than the interaction of the material with the task. For example, Holly strategically wrote and self-assessed practice essays, refining her position each time. This approach spilled into her Interview 2 essay, precluding her engagement with the quotation and question as she replicated the first few paragraphs. As a result, her essay

may seem 'downloaded' to an examiner, and not be valued as an sign of self-regulated learning, even though she set goals and chose to self-assess her writing, using her judgements to develop and refine her argument.

The value of the theory of action is that it allows for the different facets and processes of self-regulated writing that lead to such adaptations to be illuminated. It thus exemplifies and extends Sadler's (1989, p.143) idea of 'intelligent self-monitoring'. Figure 20, below, updates the theory of action, and the following sections review and evaluate the aspects key to the research questions. For eight of the ten case study writers, peer assessment updated their evaluative expertise, and so they chose to use this to prioritise and make adaptations to their writing by shifting metacognitive standards. As the importance of self-efficacy in the Oxford context became apparent part-way through, self-efficacy was not included on the first version of the theory of action, but it is present here as there is evidence that self-efficacy beliefs changed across the group, and in individual cases. Various other conditions which were not included in the original theory of action have also become visibly important.

Firstly, in yellow are personal conditions that both contributed to the self-regulation of writing and were updated by experience and by workshop learning. These conditions were shaped or even determined by task conditions, discussed next. For instance, time influenced the amount and nature of exam preparation, and in turn, writing. Other genres of writing were important conditions, too. In Chapter 6, we saw how students drew on but also adjusted A-level and tutorial essay standards. Participants also tried to enact tutor feedback, as in Chloe's Interview 1 essay (Section 7.6). College revision classes were also influential, for example in Ben's use of his tutor's strategy to 'gloss' the quotation like any other quote (Section 7.5). Basing the research questions and design on this theory of action allowed the role of these conditions to be captured.

Figure 20: Theory of action revisited



As well as personal conditions, the original Winne and Hadwin (1998) model of SRL includes task conditions. When synthesising this with Sadler (1989, 2010) to create the theory of action for the study, I elected not to include task conditions. This was for various reasons. Firstly, on the surface, the writing task was the same on all three occasions: to produce in one hour an essay in response to an exam paper and question of the writer's choice. Secondly, including task conditions would make the theory of action more complicated and I wanted to keep the focus of the research questions focussed on peer assessment as formative.

However, three task conditions in particular rose to importance in the findings chapters, and these are in grey on the model. Firstly, in the immediate situation, the one hour time limit required students to adjust standards, and as Chapter 7 showed, task perceptions and planning were brief: under time pressure, writers needed to get writing. The second and third conditions are interlinked: the time available between the workshops and the examinations themselves – a busy five week period – and the purpose and function of the examinations.

The focus in the study is on learning about and developing writing, and so to make best use of participants' time, and ensure focus on the formative, I chose for the workshops to be comment-only for various reasons. Firstly, although Sadler (2009a) includes holistic grading in his proposal, the procedures are less well described. Secondly, limiting peer assessment to comments makes peer assessment less threatening, particularly in a competitive environment. Peer grading can be problematic (e.g. Patton, 2012; McConlogue, 2012). Thirdly, in the Oxford context, the marking and classification criteria are written for examiners, and apply to the entire degree course. To non-insiders, including myself, they do not give convey clear thresholds for distinct academic standards, and so debating the differences between marks would not be meaningful as the difference between, say, a 65 and a 66, doesn't convey much information (Boud, 2017). With complex writing like Prelims essays, disagreements over marks

are inevitable, and time used debating the technicalities of the classification criteria would be better spent using examples to really draw out what is important to writing quality at the Prelims stage of the degree. Nonetheless, grades serve to confirm or contest qualitative judgements, and four respondents to the workshops evaluation questionnaire stated they would have liked to see marks, although one specifically welcomed the comment-only focus.

Although the workshops were purely formative, the essays were essentially rehearsals for the 'real' summative assessment, and so the examinations were inescapably influential. On the one hand, the examinations provided an extrinsic motivation to attend the workshops, and to adapt writing – students were aware that they needed to go beyond A-level. The examination itself is a demanding task worth teaching towards – it requires students to prepare by synthesising authors into topics, and then apply this knowledge in novel situations, which is valuable learning (Sadler, 2007; Salomon and Perkins, 1989). Nonetheless, the time available to prepare for the examinations, and the function of performance as 'proving your place', as Chloe put it, inhibited adaptation. At Interview 2, days before the examinations, many students were not yet 'at the writing stage' of preparation, to use Heather's expression. Students relied on strategies brought from A-level, and the most commonly reported preparation technique was memorising quotations. Memory is important to learning – without quotations, 'I wouldn't have had anything to say' as Ben put it at Interview 1. But for what purpose? As Darlington (2014, p.343) found in her study of mathematics students at Oxford, there is a difference between memorising to 'do' the subject – in this case, analysis and argument – and memorising to reproduce ready-made solutions or arguments, as Holly appeared to do so in her Interview 2 essay. These issues are discussed further in connection with task perceptions.

9.3 Evaluative expertise

A contribution of the study is to offer insights into the nature of evaluative expertise, and to trace the connection with writing development. The theory of action posited that participating in peer assessment workshops would enhance students' evaluative expertise, and so the metacognitive standards they would use to self-regulate writing. As noted in Chapter 3, evaluative expertise is ill-defined, and so I used my own definition: rich knowledge about quality and the salience of criteria which enables students to make qualitative judgements on work. Evaluative expertise was not directly named in the research questions. Instead, to make the research workable, as a proxy, I examined the patterns in students' comments – the criteria underlying evaluations of their own work and others'. Research Question 1 asked: During the workshops, which criteria do students use to evaluate written essay products? There can be an infinite number of criteria for writing (Sadler, 1989), and so mapping the criteria that students invoked in their evaluations of writing was important for identifying both the breadth of evaluative expertise, and the extent of convergence between student and faculty criteria. By using inductive coding rather than an existing frame, I was able to show writing quality from the students' perspective, and this turned out to be both rich and involve latent criteria which are not visible in expert descriptions of quality.

Among the broad range of thirty-four criteria codes, there was a core of criteria which were important for achieving task compliance (Sadler, 2010). *Question*, for example, was the most frequent code, and central to the web of criteria that made up writing quality, and task perceptions are discussed later. A risk of not predetermining criteria is that participants could bring their own idiosyncratic values and misconceptions to bear on assessment, but as Ben and Tessa's cases showed, discussions challenged and added to students' own judgements. The good overlap between the criteria codes and the Faculty criteria makes a validity argument for

the process (Stobart, 2012). The criteria codes identified in student comments went beyond congruence, highlighting important procedural aspects of writing, discussed in the section on standards below.

Student comments also reflected the complexity of writing quality. Care was taken during the coding process to ensure that similar codes were distinct and separate, and that segments were not coded at close codes (Section 4.11 & Appendix I). And yet students' evaluations often drew on multiple criteria at once, as the visualisation of the connections between criteria showed. To comment on argument is often to comment on the handling of the question, for example; it is not meaningful to separate out these judgments. Individual criteria cannot stand for quality; rather quality is a unitary construct (Sadler, 2010). Holistic assessment without predetermined criteria allowed students to build their own 'general impression' of each essay, as Tessa put it, and notice what was most important to quality in that particular case. In interviews, there was a sense that reading others' work made it apparent that unlike A-level, there was no fixed formula or recipe to essay writing. In Section 6.12, Tessa described how 'it's not that there are certain things the essays do or don't do and that will give them a certain mark, it's more just a general impression that you get from something'. In other words, the holistic impression of an essay may transcend the presence or absence of a particular criterion: including a specific criterion is no guarantee of 'a certain mark'. Criteria are not always manifest, or even crucial to a piece of writing. Discussing her own writing, Tessa considered how 'when you're writing to just a tick a box… it doesn't necessarily show engagement.' For Tessa, engagement emerges from an overall impression, and cannot be achieved simply by ticking criterion boxes.

As the analysis in Section 5.3 showed, 'engagement' is a keyword in the Faculty criteria. The Faculty criteria are not absent from students' judgements, then, but renegotiated and made

more feasible. Sadler's distinction between 'fuzzy' and 'sharp' criteria is helpful for examining how this occurs. In defining fuzzy criteria, Sadler (1989, p.124) gives 'originality', also key disciplinary value included in the Faculty criteria, as an example:

> 'Fuzzy criteria are characterised by a continuous gradation from one state to another. *Originality*, as applied to an essay, is an example of a fuzzy criterion because everything between wholly unoriginal and wholly original is possible' (original emphasis).

To a first-year, 'originality' may initially seem like a sharp and so unattainable criterion, particularly when the other texts about literature that students read are part of the well-established 'mass critical corpus', as Phoenix put it in Interview 1. Such publications are carefully crafted by academics; they are not a realistic account of what students could or should be expected to do in the first year of university, or indeed in timed conditions. Pre-workshops, this ambitious, aspirational sharp criterion made Phoenix 'quite worried', but through assessing essays, he was able to see Sadler's 'continuous gradation': originality in the context of peer work is different to an established academic's 'brilliant new thesis'. By Interview 2, originality became fuzzy; repurposed from 'groundbreaking', it became 'just grappling' with ideas, and interpreting them 'in your own view' (Section 6.4). Indeed, one of Phoenix's reviewers during the workshops praised his 'innovative choice' of texts, a comment which could be seen as on the spectrum of fuzzy originality.

The criterion code *originality* was, nonetheless, low frequency in the student data (Section 5.2). While every essay will be positioned somewhere on the spectrum of originality, from 'wholly original to wholly unoriginal' (Sadler, 1989, p.124), the criterion is not manifest in many essays as it doesn't arise as a determinant of quality. Recognising the conditionality of criteria – to use Sadler's (1989) term, the metacriteria for criteria – is important to evaluative expertise. At the end of Workshop 1 onwards, the workshop groups developed their own lists

of criteria, which participants perceived as more usable than Faculty criteria. However, just as some examiners do not refer to criteria while marking (Bloxham, Boyd, and Orr, 2011), students trusted their evaluative expertise, and did not see every criterion as essential for every essay. For example, Megan reflected in Interview 2 on the synthesis of criteria:

> 'It's a very good criteria [sic] but also [pause] like [pause] err it's something that like isn't blanket for every essay, maybe? […] obviously it depends on which essay you're writing as to whether you can, or whether you need to 'balance one-off examples, close reading episodes with [not] too much plot', like how… that balance shifts'.

Criteria are not 'blanket', then; when writing, they are contingent on the material – this criterion is manifest when writing about novels, for example, but is likely to be irrelevant when dealing with poetry. Amber, taking a more critical perspective, found the criteria 'broadly helpful' but 'didn't… take it too seriously because it incorporates a lot of different approaches and it feels like impossible to satisfy a kind of portion of the criteria.' She felt that assessing and discussing the example essays was more 'constructive' as:

> 'You can kind of decide what works for you, but also have a knowledge of other ways of doing it. Rather than kind of having this mass of criteria in front of you, having to pick out… umm.. *you know* which to implement and which can be left behind' (my emphasis).

Assessing the examples then, allowed her to develop her own evaluative expertise – so that '*you know* which [criteria] to implement'. This evaluative expertise is more helpful than making sense of a 'mass of criteria' – it makes it possible to prioritise and weigh up which are most important to quality. However, not predetermining criteria does raise the risk of perpetuating writers' blind spots about quality, if they have particularly tightly held beliefs and aren't willing to open these up. For example, *critics and theory* was key to the web of criteria in Chapter 5, yet Amber was alone among the case study participants in using little or even no critical

perspective in her essays, despite receiving feedback to this effect. With limited time, for Amber, using primary texts outweighed all other considerations. Holistic assessment can confirm students' existing views of quality, then, perhaps empowering resistance to criteria that are valued by others. But peer assessment is not and should not be students' only source of knowledge about quality – perhaps receiving feedback from a more authoritative source, for example a tutor, might change Amber's mind about this. Alternatively, Amber could be mirroring a tutor's preference for close analysis rather than theoretical readings. Engaging with literary theory is a particular challenge at transition to university English (Eaglestone, 2007) and students can also struggle with balancing critics' views with their own voice. At Interview 2, Amber describes 'the best [essays] as authoritative, so that they show that the candidate has read as widely as possible'. Drawing on critics could compromise her own authority, constructed through close knowledge of her primary texts. This study was limited by its exclusive focus on the first year, so it would be fruitful for future studies to longitudinally capture how students come to negotiate more ambitious criteria, and resolve tensions between critics, originality and voice.

9.4 Shifting standards

In discussing evaluative expertise, and how '*you know* which [criteria] to implement', to use Amber's phrasing, the discussion has already begun to address standards, a key contribution of this study. Metacognitive research tends to use external standards and experimental settings, and yet as Hadwin and Webster (2013) point out, real world tasks such as exam essays involve complex judgements which rely on self-set standards. This thesis contributes by giving insights into the nature of the standards students use to guide writing, and how they renegotiate long-held standards, for example, from school writing. Research Question 2 asked, 'which criteria

do students use as standards to self-regulate their essay writing processes?' Two key findings emerged. First, students often adopted procedural criteria which were latent in the Faculty list as standards for their own writing. Second, standards tended to become more reader-oriented.

Assessing essays enabled students to notice criteria which were latent in the Faculty criteria, as the visualisation showed (Figure 12, Section 5.3). Of these, *signposting* and *introduction* in particular were more frequently used as standards at Interview 2 than Interview 1. Whereas the Faculty criteria represent the examiner's perspective, students are the producers of the genre. As it is 'their' genre, the process of writing may be more apparent to students, and so may raise metacognitive awareness (Hacker et al., 2009). If writers themselves wrestle with introduction writing, introductions may be more visible. Peer texts also provide possible alternatives to improve one's own writing. Whereas the Faculty criteria describe quality after three years of study, the criteria codes that only students notice reflect their developmental stage – for example, adapting the moves in their introductions, as we saw in Tessa's case.

A key shift in standards in the pre- and post-workshops process data was towards reader-oriented standards such as clarity and signposting. Clarity is present in Faculty criteria, but signposting was one of the codes only found in student data. During the workshops, students often experienced confusion when reading others' essays. On re-reading their own essays after assessing peer essays, many realised that they did not come across as clearly as intended. As Felicia put it, 'it is not always very explicit in what I'm trying to say'. Noticing a gap between intended meaning and readerly interpretation seems to have prompted students to focus more on strategies such as signposting. This is important to writing development: Kellogg (2008) proposes that after learning knowledge transforming – noticing how well the one has translated one's ideas into written text – comes knowledge crafting. At this expert stage of

writing development, the writer is able to also juggle the representation a reader would make of the text. Kellogg (2008) theorises that maturation and working memory are important, but Lindgren et al. (2011) found that genre knowledge was also central. Thus confusion when reading others' texts prompts a focus on reader-oriented standards in one's own, and the genre moves that assist meaning and argument, such as signposting the argument in the introduction. Kellogg (2008, p.9) highlights that a general awareness of a 'fictional' reader is very different to 'being able to read the text as it is currently written from another person's perspective'. In the workshops, students cycled between reading and assessing others' work, writing and reflecting on their own writing. It seems that this combination of activities enabled students to go beyond an imagined, fictional reader, and understand how their work might be construed by others. Often in higher education by default it is predominantly or solely teachers and examiners who read work to evaluate it, and so judge quality, but peer assessment gives students experience of this role, and so by developing their reading and noticing, enables them to develop writing. A limitation of this study is that it did not include textual analysis of essays, which could have facilitated a more direct examination of the connection between peer assessment and developmental stages such as knowledge crafting. There has been great interest in Kellogg's (2008) theory, and future research should examine more closely the role of peer assessment and other interventions in knowledge crafting.

That Faculty criteria do not explicitly include student-noticed criteria which are important to writing development does not imply that such criteria should be included in official lists. Stipulating that an essay needs to have a well-structured introduction that outlines the argument would be overly prescriptive. Students need to see examples, and a range of quality for this criteria to become visible for them, and to relate it to their own writing. In addition, criteria such as *introduction* and *signposting* may not need to be intentionally invoked as

standards after a certain time. Sadler argues that holistic assessment makes criteria manifest, but these become latent again as students incorporate them into their evaluative expertise:

> 'The art of formative assessment is to generate an efficient and partly reversible progression in which criteria are translated for the student's benefit from latent to manifest and back to latent again. The aim is to work towards ultimate submergence of many of the routine criteria once they are so obviously taken for granted that they need no longer be stated explicitly' (Sadler, 1989, p.134).

One criterion for which this occurred was *range*, which was a frequent criterion code during workshops, but decreased in use as a standard between Interview 1 and Interview 2 (Table 17, Section 6.5). The examination papers each test a century of literature; thus *range* is a key difference between A-level and degree. As Heather put it at Interview 2, at Prelims:

> 'There's not even any set authors. There's literally just a set time period and you can write about anything within that stretch. Which in some ways is really amazing […] but it's also obviously a bit more difficult. And I think it's hard to go into as much depth.'

Heather commented on *range* throughout the workshops. Examples of additions made during discussions include noticing Example Essay 4 showed 'lack of range', with two novels, whereas Example Essay 5 had 'too many texts → Not fully discussed'. Here, *range* co-occurs with and makes manifest *depth*; too much of one limits the other. Across the group, as students worked out how much *range* was enough, and also feasible within an hour, it became less of a concern, and so was invoked less frequently as a standard at Interview 2. For instance, Chloe reduced the number of texts she covered in her essay, enabling her to finish the Interview 2 essay. For Heather, the challenge of *range* shifted slightly as her task perceptions began to encompass the three hour exam, not just the one hour essay. Rather than the conflict with depth, Heather was concerned with dividing material so there was *range* across all three essays, and no repetition; in dealing with five novels, her essay may not have left sufficient

material for other questions. For her, then, *range* was not 'submerged', but it was differently manifest. The nature of criteria-as-standards shift, and the implications cannot be fully captured in a fixed, static list.

Aside from criterion-based standards, process standards were also important to the self-regulation of writing. As we saw in Chapter 6, Tessa's standards for planning shifted from planning to generate ideas, to planning for structure and finally planning for the introduction. Emotion arose in the dataset more than anticipated, and a limitation of the theory of action is that it does not take sufficient account of this, partly because it is based on the Winne and Hadwin (1998, 2008, p.297) model, which sees affect as 'weaving' through self-regulated learning but does not give it prominence as a discrete aspect. Metacognitive feelings may be fleeting (Efklides and Volet, 2005), but they can also shape decisions, as we saw with Chloe's choices around 'scary' question-quotations in Chapter 7. Nonetheless, by Interview 2, Chloe was able to finish her essay in the time, a more pressing priority, and progress from her January collections.

Metacognition and emotion can be combined in standards, which makes it harder to change. At Interview 1, Phoenix used a familiar routine – starting the essay by analysing a 'simple' quotation, rather than the title quotation – in order to 'get into… the rhythm of writing' (Section 6.3). The standards at play here intertwine the emotional and (meta)cognitive. Using this quotation to write about 'stuff that we've been doing since like aged 12' both enables him to feel comfortable and do something familiar which he knows he can do well. Participants very much wanted to write about their ideas and 'juicy quotations', as Phoenix put it, and to be rewarded for doing so, exams ultimately demonstrating knowledge. Like Chloe and Ben, Phoenix initially saw the quotation-question as a threat to his argument. However, he did adapt, coming to see ideas and their construction as complementary: 'now I'm focussing more

on how I write not just what I write'. Through the workshops, Phoenix saw the value of engaging with the quotation, making annotations, for example noticing during a discussion how an example essay writer 'makes the quotation personal'. At Interview 2, he seems to have found a balance between his voice and the question:

> 'I'm a lot more focused on the question now, and… not necessarily... like chained to it but engaging it to make sure that I can make, make it seem that I'm using it to my benefit, not I'm being enslaved by the question.'

Again, there is a potential tension between what Phoenix wants to write, and what the question demands – almost a power struggle. However, Phoenix seems to have resolved this: he is in charge and 'using it to my benefit'. Thus the question is not just an externally imposed standard but one that enhances the quality of his work. The advantage of peer assessment is, then, that it allows students to see for themselves the importance of criteria such as the question/quotation. This may be more persuasive than telling them to do so, for example through feedback or assessment criteria. In discussing standards around the question, we have moved towards task perceptions, and the important debate around task compliance and creativity.

9.5 Task perceptions, task compliance and creativity

While engaging with the question-quotation may seem obvious to teachers – and to students once they have seen its contribution to quality for themselves – it is an example of important tacit knowledge that is best developed through evaluative experiences. The examiners' reports excerpted in Chapter 7 show that shaping knowledge to the question is important for task compliance and success. This is a valuable skill: a valued sign of learning is being able to transfer knowledge to new situations, but without replicating old routines inappropriately

(Bereiter and Scardamalia, 1984; Schwartz et al., 2012). Prelims essays are reading-to-write tasks (Flower et al., 1980). Using the Winne and Hadwin (1998) model of SRL as the basis for the theory of action allowed for a focus on task perceptions phase, crucial not only in this context but to writing more widely (Flower and Hayes, 1980; van den Bergh and Rijlaarsdam, 2001). Research Question 3 asked: (How) do students' task perceptions shift over the course of the study? The example from Phoenix highlights how as *question* became a more important standard over the period of the study, students also became more focussed on task perceptions. Chapter 7 showed that most pre-writing time was spent reading the question paper to decide on a question, and then planning: time dedicated to thorough reading and interpretation of the question was more limited. The chapter also showed how the 'downloaded' essay comes into being. Task perceptions are formed in the interaction between the task itself and what the writers bring to the task, and in the case of Ben's Interview 1 and Holly's Interview 2 essay, the emphasis was on retrieving and rehearsing knowledge, with the reading of the question-quotation itself narrowed to the theme that they wanted to write on. Nonetheless, Chapter 7 also showed how the workshops raised Ben's awareness of the importance of the quotation to writing quality, and how at Interview 2 he invested time deconstructing the quotation. All except Holly at Interview 2 addressed the quotation at some point. However, this tended to be in order to ensure task compliance – showing engagement with the question – rather than to enhance the quality of their argument. Task perceptions are formed in the interplay between question and reader/writer, but in some cases students' efforts were limited by the task itself. Ben put considerable effort into deconstructing the quotation, but was limited by the question's emphasis on 'moralists' – a group of writers of which he was unaware. In this case, the scaffold of the question became a confounding rather than supporting factor, and this challenge was also voiced in tutor interviews.

While the examination itself was beyond the scope of the study, and the interview guide, some participants made observations of note in Interview 2, just a week or even a few days before Prelims. For instance, Amber considered that the nature of the question contributed to her perceived success in the Interview 2 essay:

> 'I think under an hour's exam, there's always the chance that the quotations just don't jive with what you've revised […] I think this quotation brings up a lot of very interesting issues in a lot of different ways in the two authors that I've picked, they might bring a relevant but maybe less interesting or less original issues for other poets, which can be problematic because… as much as they say that they're grading on [criteria], I know that the examiners are also looking for originality, because it's two examiners reading over 200 papers and they're probably very tired.'

Amber is highly aware that she is writing for an over-worked reader, whom she needs to engage and interest, and so for her, originality transcends all other criteria; it is a prized disciplinary value. Yet the construction of originality depends on the interaction of the quotation and the material the student brings to the task, and so there is an element of 'chance'. Amber preferred the portfolio paper, which still put students under 'intellectual pressure', as 'suddenly, for the next two weeks, you're questioning your every single hour you're not spending working on the portfolio.' Interestingly, several students stated that performing under pressure did seem to be a key purpose of the exams, and acknowledged that it could be an important life skill. While Amber and others voiced a preference and enjoyment of portfolio assessment, and such coursework also gains some favour in the literature for reasons of validity (e.g. Yancey, 1998; Marshall, 2017), a drawback is that it is time-consuming. Students reported that writing the two portfolio essays was their sole focus for two weeks of the eight week term, with additional preparation in the run-up. Assessing four papers as portfolios is perhaps unrealistic in the context of short terms, then. As re-interpreting knowledge and ideas through a question-quotation seems important to learning and writing in

English, an alternative solution might be to set 15-30 minutes of dedicated reading time at the beginning of the paper. This may encourage students to engage in task perceptions strategies such as carefully selecting a question, and interpreting the question-quotation, as well as using these task perceptions to plan an answer that goes beyond retrieval. A further solution may be simplify or even remove the question scaffolds that accompany the quotations, directing student attention to the quotation itself, as in Finals.

In this study, Sadler's (2010) notion of task compliance primarily meant addressing the question and quotation: this was the central criterion in the web of criteria, from which almost all others stemmed. Other criteria were important, too, but their configurations varied, while *question* was central. The workshops intentionally employed a range of essays of differing quality, with none considered truly model. This was so that students didn't perceive a particular essay as instantiating task compliance, and so replicate it. In practice and in the literature, sharing examples gives rise to fears of plagiarism, which became real in one study, where only two exemplars were shared (Handley and Williams, 2012). In other studies, following interventions, students have used exemplars or criteria like recipes in their work (Bell et al., 2013; Bloxham and West, 2007). However, in this study, reading many and different examples mitigated against this. Ben, for instance, did not experience holistic assessment as 'conformative', as Torrance (2012) puts it. Learning about writing by 'seeing so many' examples meant that he didn't feel he 'had to copy' or replicate the 'good' essays. He emphasised that: 'I wouldn't now remember one essay and think 'I have to write like that person''. As Amber put it above, this meant that 'you can kind of decide what works for you'. Seeing multiple but not perfect examples was also important for self-efficacy, as discussed later (Section 9.8). However, Nicole expressed some tension between criteria and 'big thinking' (Section 6.12). She experienced conflict between focussing on overt standards and 'just… be[ing] original and writ[ing] what's interesting'. There seem to be two implications here. First,

that explicitly focussing on 'the criteria and… not the boring stuff, exactly, but the necessary things' is a potential brake on creativity. As discussed in 6.12, there are inevitably constraints on creativity when writing for a reader, particularly when that reader is an examiner, and there is a power imbalance. To a certain extent, for work to be valued by its readers as creative, it has to do at least the 'necessary things', which create and sustain a reader-writer relationship. Sadler (1989, p.128) argues that creative work is in fact 'highly disciplined' and becomes so 'when the producer, by being thoroughly conversant with the characteristics of the discipline or genre, understands when and how to transcend the normal boundaries.' However, writing an examined, timed task offers limited scope for this.

Second, as Nicole points out, the workshops did not include discussion of the metacriteria for when 'big thinking' and ideas outweigh other concerns (Sadler, 1989). The workshops were run outside the discipline, by a researcher – me – with only some knowledge in the subject content. This was both a strength and limitation of the design. On the one hand, a researcher-led intervention ensured fidelity of implementation, and also meant that the focus in the workshops was on writing rather than literary ideas. Students did not see me as a disciplinary expert, and so there was less risk of writing being forgotten, or 'swept aside', as Nicole might put it, in favour of literary debates. On the other hand, we did not and could not discuss the possibility of outstanding ideas over-ruling other aspects of practice. In the analysis, too, knowledge was separated from other criteria to make the coding more practicable.

However, as Nicole raises, at degree level, to what extent can knowledge and ideas and writing be separated? We construct knowledge through writing; the knowledge constituting process of writing is in particular how we discover new ideas (Galbraith, 2009). Knowledge constituting during the timed task was experienced as problematic for some students. It is risky to discover new ideas or have task perceptions shift when writing carries consequences. When we read

written products, can 'really, different interesting ideas' become the most important manifest criterion and overrule other aspects of writing quality? Having a disciplinary expert lead or co-teach the workshops might allow this question to be answered, and so the workshops would ideally be run within the discipline. Yet there is unlikely to be an easy or definitive answer: English privileges different interpretations and readings; as participants recognised, different readers notice and value different aspects of writing, even when sharing core criteria.

9.6 The theoretical problem revisited

This study set out to address a theoretical problem. The wider field of assessment is preoccupied with outcomes, as the ongoing debate around the overall effectiveness of Assessment for Learning shows (e.g. Kingston and Nash, 2011; 2015; Briggs et al., 2012). This is understandable – measuring learning is a central activity and concern of the field, and arguments for raising achievement are important, not least to policy makers (Black and Wiliam, 2003; Baird et al., 2017). However, in order to develop formative assessment theory and so develop effective practices, we need an understanding of processes and of domain-specific learning (Bennett, 2011). This thesis contributes by using a theory of action that brings together formative assessment, self-regulated learning and writing.

The study has shown the importance of evaluative expertise to writing. In this study, I used a working definition of evaluative expertise as rich knowledge about quality and the salience of criteria which enables students to make qualitative judgements on work. The analysis has shown that students indeed draw on a wide range of criteria, which are interconnected and also are broadly congruent with Faculty criteria. Knowledge of metacriteria is also central to evaluative expertise, both in terms of which criteria are non-negotiable and so essential for

task compliance. Evaluative expertise is not fixed: it is dependent on the essay and situation. The theory of action also made it possible to examine how developing evaluative expertise influenced students' own standards and task perceptions. This study gives insights into how students self-regulate during tasks, and also how their self-regulation develops over time.

In the time since I designed the theory of action and study, research interest in evaluative expertise has continued to grow. For instance, Tai, Ajjawi, Boud, Dawson and Panadero (2017) have drawn together a helpful overview of the area and identify how a range of activities can be leveraged to promote the development of evaluative expertise. They put forward 'evaluative judgement', a term also used by Nicol (2013), and provide a clear definition: 'evaluative judgement is the capability to make decisions about the quality of work of self and others' (Tai et al., 2017, p.5). However, I prefer to use 'evaluative expertise' for several reasons. First, 'evaluative judgment' appears tautologous: 'evaluation' and 'judgment' are often used interchangeably, although perhaps erroneously. In addition, 'evaluative expertise' emphasises that students can be experts in learning, too, and share in Sadler's (1989) 'guild knowledge', prized by academics and professionals. It positions students as on their way to becoming expert, rather than being only subjects of assessment. The use of 'expertise' also reflects the emphasis in the SRL field on students becoming researchers and experts in their own learning (e.g. Winne, 2005).

A further reason to prefer the term 'evaluative expertise' is that students already make judgments and decisions about their work. Researchers and teachers often overlook the fact that students self-regulate from a very young age (e.g. Perry, 1998); teaching is not always set up to allow students to do so (Vermunt and Verloop, 1999). Also, students may not self-regulate very well, or they may adopt goals that we as teachers would not hold as ideal, as in the case of Holly. Anyone can make decisions about the quality of work, particularly if using a

rubric or proforma, where just a tick or number is required. The term 'evaluative expertise' reflects the richness and complexity of the judgements made by students in this study. Holistic assessment without predetermined criteria was challenging. As Claire put it in her evaluation questionnaire, 'it makes you think how an essay is good rather than just whether it is or not'. In students' future professions, judgements and decisions about quality are also likely to be complex and emerge through discussion and negotiation with colleagues. As Sadler (2010, p. 547) puts it, 'in the world beyond graduation, hosts of things of all descriptions have to be appraised for a variety of reasons and purposes, and to seek a rubric for each is unrealistic.' It is important to consider the sustainability of assessment practices: the extent to which assessment 'meets the needs of the present without compromising the ability of students to meet their own future learning needs' (Boud, 2000, p.151; Boud and Soler, 2016). It is beyond the scope of this discussion to consider the sustainability of the three-hour exam as an assessment format, but holistic peer assessment appears to hold promise as sustainable assessment. By granting students both the autonomy and responsibility to define their own criteria and standards, students have to engage with texts for what they are, and weigh up what contributes to quality. The Oxford essay is a very particular genre, but the skill of judging quality without pre-determined criteria is highly applicable and transferable to novel products. While some of the codes that emerged from students' noticing are bounded by the discipline and context, most are important for writing quality and communication more broadly.

Capturing and measuring evaluative expertise is a challenge for the field; this is perhaps why many studies continue to compare students and teacher ratings (e.g. Moxley and Eubanks, 2016). An obvious limitation of this study is that it is small-scale in terms of participants, but this enabled me to generate and collect in-depth and longitudinal data. Although analysing qualitative judgements is labour intensive, future studies should take up this challenge. Researchers are already using automated content analysis to parse comments (Anson and

Anson, 2017), although further advances in computational linguistics are required for more nuanced analysis. A further challenge which I did not address was measuring the quality of students' evaluative expertise. I could have perhaps compared responses and annotations made pre-discussion to tutor interviews. However, only two tutors reviewed each essay; as judgement is divergent in English and writing, more raters would be needed. Comparing qualitative judgements is complex, not least because of the slippery nature of academic terminology. If both comment on the argument, how can we know that they share the same meaning of argument? In addition, each response inevitably draws on previous judgements – as Wyatt-Smith and Klenowski (2013, p.37) put it, judgement is 'cumulative'. This makes it extremely difficult pre- and post-test meaningfully: aside from the issue of sourcing comparable essays, judgements are not independent.

A contribution of this study is, then, that it acknowledges and represents the complexity of evaluative expertise, and captures traces of the processes through which it is developed. However, using annotations and additions made in discussion on example essays is not a failproof method – in some instances, students forgot to switch pen during the discussion, and inevitably, some students made more copious annotations and additions than others. Future research should further investigate how evaluative expertise is developed, using methods such as stimulated recall from video recording of discussions to capture the role of talk. Research should also compare different modes of assessment to evaluate discipline-specific suitability. To judge the sustainability of such practices (Boud and Soler, 2016), studies should also consider whether and how evaluative expertise carries within and beyond the university context, and how transfer can be supported and facilitated.

9.7 Evaluative expertise and the role of feedback

Research Question 4 asked: (How) does peer feedback contribute to students' self-regulation of writing? In some ways, it should be surprising that only two case study participants had directly implemented peer feedback suggestions by the Interview 2 essay, four or five weeks after the workshops. Students received a mean of 496 words of comments, which drew on a core of criteria used throughout the workshops, and offered a good proportion of solutions as well as pointing out praiseworthy and problematic features. Nonetheless, the rate of uptake was low, and even lower than in other research, where students were then given the opportunity to revise the assignments before grading (van den Berg et al., 2006; Walker, 2015; Patchan et al., 2016). For reasons of ecological validity, students did not revise and resubmit essays to this study, and so had free choice on use of feedback. Winne and Hadwin (2008, p.313) consider that 'definitionally, self-regulation is a phenomenon by which options are considered after a product has been created'. Peer feedback added to the options that students were already weighing up. A contribution of this study is to bring to light why students didn't necessarily enact peer-offered options directly, and the uses to which they put feedback instead.

Firstly, analysis confirms that as self-regulating writers, students 'filter' feedback through their own beliefs and metacognitive knowledge of writing (Butler and Winne, 1995). As described above, Amber prized detailed knowledge of primary texts over critics and theory, and so didn't act on feedback as suggested, trusting her own evaluative expertise to guide her priorities. Also, from a recipient's perspective, feedback may not be as straightforward to enact as others might assume. Phoenix wanted to act on his three reviewers' recommendations to write a more developed conclusion, but first needed to learn to cut back on 'some of my lovely points' first in order to leave room. His case exemplified how students may experience conflicting

standards, in this case, a tug-of-war between 'fit[ing] in as much as I can' and writing a conclusion to meet reader expectations. It is difficult to unhook from long-held standards when they are rooted in enjoyment of subject knowledge, and presumably have previously brought success.

In other cases, students had already predicted a solution that reviewers then also offered, as in the case of Nicole deciding to move through points more quickly. Does such a high level of awareness and commitment to personal standards make feedback redundant, then? The data implies not. In responses to the workshops evaluation survey, the majority of participants – 13 of 21 – considered that reviewing and receiving feedback were equally useful. Writing is fundamentally a communicative act. Students want a reader; essays are produced to be read. Often the emphasis in formative assessment is on closing gaps (e.g. Sadler, 1989), on next steps (e.g. Black and Wiliam, 2009). However, taking stock and noticing the strengths of work, is valuable, too, particularly for self-efficacy beliefs, as discussed in the next section. For time-pressured students, knowing what to keep is important. For instance, feedback confirmed the success of Tessa and Heather's new introduction moves, which they then carried through to the Interview 2 essay. Feedback provided Phoenix with 'verification' that his choice of texts and topic was apt.

These findings exemplify the wider arguments in the literature that as students should be recognised as self-regulating learners, so conceptualisations of feedback and formative assessment must change. The focus needs to move beyond feedback as simply authoritative 'telling' (Sadler, 2010, p.), or as 'individual acts of information provision and reception' (Boud and Molloy, 2013, p. 706). A theoretical notion that has become influential in practice is the argument made by Sadler (1989), drawing on Ramaprasad (1983), that feedback only becomes feedback when it is acted upon; when it is used to close the gap between current work and the

aspired-to goal. Perhaps this influential view, and the compelling research evidence for feedback in reviews such as Hattie and Timperley (2007) has led to the orthodoxy that *all* feedback should always be implemented. Such an assumption has led to schools instigating practice like triple marking, where in order to evidence that marking is promoting progress, students respond to feedback, and their responses are then marked again, even though there is no research evidence for these particular practices (Elliott et al., 2016). It has also led to an emphasis in higher education policy on the quality of teacher feedback (Nash and Winstone, 2017). Feedback is undeniably important. However, it is just one component or strategy of formative assessment, as Black and Wiliam (2009) highlight, for example. Their definition of formative assessment involves the use of assessment evidence in decision making about 'next steps'. Such decisions 'are likely to be better, or better founded, than the decisions [teachers, students] would have taken in the absence of the evidence that was elicited' (ibid, p.9). There is no cast-iron guarantee that these next steps will improve learning. In this study, students' decisions about what to adapt next were not always ideal decisions, perhaps. But in many cases, combining peer feedback with their own judgements meant decisions were 'better founded'. Feedback, then, *was* used, but students had already spent a considerable amount of time developing and refining their own judgements about quality and their own writing, it was not the only source. Rather, it became additional evidence that was 'amalgamated' with their own judgements, to use Ben's term.

The framing, set-up and uses of feedback in this study then, align with what Boud and Molloy (2013) term Feedback Mark 2, where feedback is centred around students as proactive, agentic self-regulating learners rather than teachers or authoritative others generating and perfecting messages for improvement. In this view of feedback, students are also 'elicitors of knowledge for improvement' (Boud and Molloy, 2013, p.705). This entails dialogue and interaction with others, also central to other contemporary conceptualisations of feedback in

higher education (e.g. Nicol, 2010, Carless et al., 2011). Although the Oxford system affords students more opportunities for face-to-face dialogue and feedback than other contexts, the use of anonymous peer assessment in this study precluded any immediate face-to-face dialogue between reviewers and writers. Coding of feedback showed that some participants took a dialogic stance anyway, questioning writers, or making explicit their own assumptions and interpretations. Students could also annotate their feedback, for example thanking reviewers ('aww cheers') or questioning the reviewer's interpretation ('maybe I am wrong but I thought the quote with the question fused the two' – Megan). In the context of perceived exam pressures, and institutional policies of anonymity and privacy for Prelims and collections marking, it was important for peer assessment to be double-blind, too. Anonymity made it easier to be 'honest' (Tessa, Interview 2), although a small number of reviewers tried too earnestly to be honest, to the point of being unnecessarily 'blunt', as Rachael and Ben found in Section 8.7. In this context, making peer assessment non-anonymous and face-to-face to provide further dialogue would demand a lot from students; it can take time for sufficient trust to be developed (Carless, 2009; Carless, 2013a). Nonetheless, if such opportunities were built into the curriculum, as many higher education assessment theorists suggest (e.g. Sadler, 2010; Boud and Falchikov, 2007; Boud and Molloy, 2013), it would be possible to use various modalities of peer assessment, building towards non-anonymous. One recent study has showed that this is possible, although the scope was limited to one four-week group assignment (Rotsaert, Panadero and Schellens, 2018). Further research is needed to investigate how peer assessment practices can create dialogue while limiting emotional demands and risks, and how peer assessment practices can be structured across an entire degree course.

9.8 Self-efficacy

The importance of self-efficacy in the context of exam writing became apparent in the pre-workshops data collection with both tutors and students. Thus, Research Question 5 was added to the study, asking: 'what is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing?'

In the wider literature, self-efficacy mediates strategy use and is correlated with achievement, but any connection between the two relationships has yet to be established (Bruning and Kauffman, 2015). This research has focussed on the impact of peer assessment on self-efficacy beliefs, but it also showed that self-efficacy played a role in task perceptions, for example in whether students tackled the 'scary' open quotation items (Chloe, Interview 1). Findings from the questionnaire designed for the study suggested that students became more confident about tackling the question-quotation. Students' self-reports increased in score on most items between the beginning of Workshop 1 and the end of Workshop 4, and in particular on items related to expectations, quality, and handling the question (Section 5.9). To make self-efficacy judgements, students draw on their knowledge of task demands (Bandura, 1997). By enabling students to develop a more realistic understanding of the task vicariously, by reading and assessing other's writing, the daunting task of writing under pressure seem more 'achievable', to use Megan's term. As she commented in Interview 2:

> 'Sometimes you think kind of think like, being at Oxford or whatever that everyone's writing these brilliant essays all the time, and so like having the confidence to kind of like say "oh actually, it's like a good essay, but it's still got flaws, too" like makes you think, "well, my essay's also got some flaws, but it's got some good bits as well" ***so*** I could kind of write this, if that makes sense.'

The 'so' here is my emphasis, and is important: as a result seeing what others have achieved, Megan is able to more realistically self-evaluate her work and so her own capacity to write a

good essay. The workshops allowed students to take stock of the strengths of their work, and genuinely positive feedback provided a boost for some students (Section 8.9).

These findings resonate with the wider research, which points to both how the transition to university is associated with a confidence dip (Putwain and Sander, 2014) and the importance of self-efficacy for belonging. Compared to course enrolment, the workshops attracted a relatively high proportion of women and state-school educated students. In Reay et al.'s (2009, p.1113) interview research with working-class students attending an elite university, assessment held an important but delayed message for students: 'until I did well in the first year, I just didn't know whether I was clever enough to be here'. This resonates strongly with Chloe's comments in Interview 2 that Prelims test

> 'All the sort of reasons, they're like, that's why you should have been let in here [laughter] I guess, just sort prove your place slightly, I think [laughs]'

For this reason, Prelims exams are stressful. In Interview 1, she voiced that the exams:

> '…show… how well you can do, under pressure, I guess. Because I mean anyone, probably anyone could write a good essay if they're given enough time.'

Prelims results, then, give confirmation as to whether 'you should have been let in here' – examinations have discriminatory power and as with school assessment, provide labels and influence identity (Reay and Wiliam, 1999). A valuable contribution of the workshops, then, seems to have been to open up others' writing for scrutiny, and allow students to see that they fit in, that their writing is similar. Research, particularly at elite universities such as Oxford, should consider further the role of peer assessment and self-efficacy in transition and in developing identity. Students' pre-university writing experiences vary considerably, and

interventions that demystify success and foster self-efficacy could make the transition to such environments smoother and, importantly, more inclusive.

The questionnaire findings were more positive than for extant research on rubric-referenced self-assessment, perhaps because these studies used only one exemplar (Andrade et al., 2009; Covill, 2012) or even none at all (Panadero et al., 2013). In this study, it seems using a wide range of examples enabled students to both see the task as more manageable and judge the quality of their own work. A further strength was that the workshops provided multiple possible sources of self-efficacy: example essays (vicarious), feedback (social persuasions), the combination potentially prompting re-appraisal of one's own essay as a mastery experience (Bandura, 1997). However, a limitation is that self-efficacy was only introduced as a focus part-way through; longitudinal data would have enriched these findings. Also, it is important to examine the longer-term ebb and flow of self-efficacy: a singular success can be discounted as a one-off, as Chloe almost appeared to do in Section 8.9 (Bandura, 1997). Future research should consider in more detail the variations of peer and self-assessment that contribute to self-efficacy, and how self-efficacy fluctuates and develops over the longer term.

9.9 The practical problem revisited

The thesis has also addressed a practical problem: the lack of writing instruction provision for the majority of students transitioning to university. Peer assessment was posited as a way of embedding learning about writing into the discipline. This study has shown that holistic peer assessment is practicable, and a valid method of developing awareness of quality. Peer assessment did support writing development – eight of the ten case study participants made adaptations linked to workshop learning. The peer assessment intervention shared some

similarities with genre pedagogies, namely in-depth discussion of texts, down to sentence level (Swales, 1990; Flowerdew, 2015). A key site of genre-related adaptation was the introduction. Six writers made adaptations, cutting general contextualising moves in favour of getting straight to the question or topic, and/or adding moves to outline their argument (Section 5.6). Analysis of Tessa's additions on an example essay introduction showed how the discussion drew her attention to actual and potential moves (Section 6.10). A limitation of genre pedagogies is that they tend to use model texts rather than authentic performances of genre (Devitt, 2015). The advantage of peer assessment is that students gain exposure to a variety of examples and so can evaluate which moves work and are also feasible within the time.

However, peer assessment is not a standalone writing pedagogy. Assessing example essays may develop students' declarative knowledge of quality and genre, but the picture is less clear for procedural, strategic knowledge. The self-efficacy questionnaire results mirrored this potential gap between the declarative and procedural metacognitive knowledge of writing. The largest and most highly statistically significant differences between Workshops 1 and 4 were for items to do with quality, conventions and expectations, writing an essay that meets such expectations, and understanding and handling the question (Section 5.9). For two items on planning for organisation and for generating ideas, however, the group median scores remained the same. Sadler (1989, p.138) acknowledges:

> 'The ability to evaluate others' or one's own work is not necessarily matched by the ability to produce… the possession of evaluative expertise is a necessary (but not sufficient) condition for improvement.'

During the design of the study, I was aware of the potential for a declarative-procedural gap, and it cemented several design decisions. Firstly, writing in a new genre requires practice (Kellogg, 2008; Artemeva and Fox, 2010), and so students wrote under timed conditions.

Secondly, in Workshop 4 I provided and we discussed a handout that alongside the synthesis of criteria offered writing strategies, ranging from deliberate practice of introductions and conclusions (Kellogg, 2006) to freewriting to develop ideas during revision (Elbow, 1998). However, few students had had time to try ideas from the handout by Interview 2. An alternative source was college revision classes, and Ben and Megan in particular tried out strategies for task perceptions and planning respectively. While peer assessment can develop students' evaluative expertise, then, for productive expertise students need practice and strategy suggestions. This is particularly important during the first year of university, otherwise students will have to rely on school genre knowledge and standards (Bawarshi and Reiff, 2011; Baker, 2016). For example, at Interview 1, Megan 'planned in quite an A-Level way'. By Interview 2, by combining workshop learning and a tutor's strategy, she was able to move beyond this 'blocked-out' structure. A strength of the research design is that is allowed the influence of other teaching inputs on writing to be captured.

Holistic peer assessment was well suited to both the institutional context and the subject discipline, which presented an ideal case of sorts. It may be harder to enact in other situations, where assessment criteria are more prominent, students are used to using rubrics, or the institution requires their use. Nevertheless, holistic and analytic assessment are not necessarily mutually exclusive (Sadler, 1989). Another possibility might be to have students first holistically assess exemplars, noticing what stands out as important to quality, later introducing a rubric. This could enable discussion of criteria beyond those sampled by the rubric, making important aspects of quality visible. Rubrics can be helpful artefacts – they provide a far more succinct summary of quality than a tangled web of criteria. They also often state levels of achievement, which this study did not address. However, simply providing a rubric or briefly practising is unlikely to give rise to an appreciation of latent criteria which

might be helpful to writing. Future research should compare rubric-led, holistic and combined approaches to peer assessment to compare the affordances for evaluative expertise.

Discussion is important to the success of formative assessment (Black, 2015), and the additions that participants made highlight the contribution of talk to learning about writing. Such discussion requires the teacher to strike a balance between eliciting and valuing students' ideas, while not giving prominence to their own views (Carless and Chan, 2017; Hawe et al., 2017). At the same time, the teacher must not 'abdicate' responsibility for 'engineering' learning through discussion (Black and Wiliam, 2009, p.22). In this study, the teaching notes I produced from tutor interviews were indispensable, and I used them to raise questions on areas which students hadn't brought up. It was rare that I had to volunteer any missed points; this tended to be around issues such as referencing. Such dialogue requires teachers to see their role as facilitating learning, rather than transmitting content (Carless, 2013b). As we know from the school literature on English, this is challenging. While teachers' espoused beliefs may align with the principles of Assessment for Learning, the structuring of classroom dialogue and practices may not afford students much autonomy (Marshall and Drummond, 2006; Dixon, Hawe and Parr, 2011). In this study, the size of workshop groups was determined by timetabling constraints, with the largest at 10 students. This made it easier for everyone to have a voice. In a different setting, with larger teaching groups, after individual annotations and responses, students could then discuss in small groups, before coming together as a class. For example, Carless and Chan (2017) had students assess two exemplars before coming to class, and do pair and group work before teacher-facilitated discussion. This study is a rare example of formative assessment dialogues within higher education, and research should focus further on observing the features of productive classroom dialogue, as Black (2015) has called for.

The time commitment to the workshops was not insignificant, at around 8 hours. However, if such interventions are embedded into degree course design, as many theorists suggest (e.g. Sadler, 2010; Tai et al., 2017), assessing examples could be spread out over a longer period. The selection of examples is crucial. A range of quality is important (Sadler, 2009a), but a possible limitation of this study was that sampling was limited to previous cohorts' collections and practice essays, rather than Prelims scripts. None of the example essays were truly outstanding, which may have contributed to Holly and Amber's decisions not to adapt. A longer-term approach to peer assessment could allow for assessing example and peer essays at several points over the year, with the range of examples carefully selected to the current level of student learning, ensuring a balance between coping examples and challenge. In this study, the limited and busy time between workshops and exams presented a barrier to adaptation; few participants wrote practice essays outside of the research sessions. Course designers should also ensure that students have sufficient time and space to adapt their learning and writing after peer assessment.

To summarise the implications for practitioners, holistic peer assessment is feasible and can enable students to notice a wide range of criteria, both congruent with and beyond manifest Faculty criteria. However, careful selection of examples and planning of discussion is required. To avoid a 'recipe' approach, multiple examples are needed (Bell et al., 2013). Anonymity and receiving multiple reviews were important to peer-to-peer feedback in this study. Various researcher-developed peer assessment platforms are available commercially (Schunn, 2016; Moxley, 2013), and many virtual learning environments also have tools, reducing the administrative burden. However, it seems that face-to-face discussion of examples first is important both to train students to give good quality feedback and develop evaluative expertise. An important accompaniment to peer assessment is discussing and providing ideas

for strategies to develop writing. In this way, writing instruction can be embedded into the discipline.

9.10 Conclusion

This study has demonstrated that peer assessment shows promise for supporting self-regulated writing development. All ten case study writers adapted their writing during the study, and for eight, these changes could be linked to workshop learning. Students had limited time to adapt their writing, but prioritised adaptations, choosing on which standards and aspects of feedback to focus. Assessing example and peer essays holistically gave rise to a wide range of criteria and made tacit, latent criteria visible. Students adjusted their standards, focussing more on procedural and reader-oriented aspects of writing. Focus on the question increased, and began to encompass the accompanying quotation. Peer assessment also had a positive impact on students' self-efficacy beliefs, particularly important during the transition to an élite, competitive institution.

On a practical level, the thesis shows that Sadler's (2009a, 2010) proposition is practicable and can serve to support learning about writing within the discipline. While peer assessment is not a standalone writing pedagogy, it enables students to develop their evaluative expertise, and to see their writing in the context of others', impacting on self-efficacy beliefs for writing. When accompanied with suggestions for writing strategies, and the time to practice and adapt, peer assessment workshops could make the transition to university writing smoother and more inclusive.

On a theoretical level, the contribution of the thesis is grounded in the way it combines insights from different literatures and traditions. The study contributes to formative

assessment theory by showing not just *whether* students learn through peer assessment, but *how.* It gives insights into the development of evaluative expertise in writing, and the role of facets of SRL such as task perceptions and metacognitive standards. The thesis also highlights the need to consider emotion and self-efficacy in assessment. While the thesis does not offer results on outcomes, the detailed, rich findings on disciplinary-specific self-regulated writing processes should be helpful to researchers and practitioners alike.

References

Adachi, C., Tai, J. and Dawson, P. (2017), 'A framework for designing, implementing, communicating and researching peer assessment', *Higher Education Research & Development.*

Adams, A.-M., Simmons, F., Willis, C. and Pawling, R. (2010), 'Undergraduate students' ability to revise text effectively: relationships with topic knowledge and working memory', *Journal of Research in Reading* 33(1), 54–76.

Alamargot, D. and Fayol, M. (2009), 'Modelling the development of written composition', in R. Beard, D. Myhill and J. Riley (eds), *Sage Handbook Of Writing Development*, London, Sage, pp. 23–47.

Alder, E. (2018), 'Becoming a student of English: Students' experiences of transition into the first year', *Arts and Humanities in Higher Education* 17(2), 185–203.

Alexander, P. (2013), 'Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration', *Learning and Instruction* 24, 1–3.

Allal, L. (2010), 'Assessment and the regulation of learning', in P. Peterson, E. Baker and B. McGaw (eds), *International Encyclopedia of Education*, Vol. 3, Oxford, Elsevier, pp. 348–352.

Anderson, T. and Shattuck, J. (2012), 'Design-based research: A decade of progress in education research?', *Educational Researcher* 41(1), 16–25.

Andrade, H. and Du, Y. (2007), 'Student responses to criteria-referenced self-assessment', *Assessment & Evaluation in Higher Education* 32(2), 159–181.

Andrade, H. L., Du, Y. and Mycek, K. (2010), 'Rubric-referenced self-assessment and middle school students' writing', *Assessment in Education: Principles, Policy & Practice* 17(2), 199–214.

Andrade, H. L., Wang, X., Du, Y. and Akawi, R. L. (2009), 'Rubric-referenced self-assessment and self-efficacy for writing', *The Journal of Educational Researc*h 102(4), 287–302.

Anson, I. G. and Anson, C. M. (2017), 'Assessing peer and instructor response to writing: A corpus analysis from an expert survey', *Assessing Writing* 33, 12 – 24.

Artemeva, N. and Fox, J. (2010), 'Awareness versus production: Probing students' antecedent genre knowledge', *Journal of Business and Technical Communication* 24(4), 476–515.

Atherton, C. (2007), 'Balancing acts: Preparing students for university in the mixed-ability classroom', *International Journal of Adolescence and Youth* 14(1), 65–76.

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

Baird, J., Andrich, D., Hopfenbeck, T, and Stobart, G. (2017), 'Assessment and learning: fields apart?', *Assessment in Education: Principles, Policy & Practice*, 24(3), 317-350.

Baker, S. (2017), 'Students' writing in transition' from A-levels to university: how assessment drives students' understandings, practices and discourses', *Assessment & Evaluation in Higher Education* 42(1), 18–36.

Ballinger, G. J. (2003), 'Bridging the gap between A level and degree: Some observations on managing the transitional stage in the study of English literature', *Arts and Humanities in Higher Education* 2(1), 99–109.

Bandura, A. (1997), *Self-Efficacy: The Exercise of Control*, New York, Freeman.

Bandura, A. (2006), 'Guide for constructing self-efficacy scales', in T. Urdan and F. Pajares, (eds), *Self-efficacy beliefs of adolescents*, Greenwich, IAP, pp. 307–337.

Bannert, M. and Mengelkamp, C. (2008), 'Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?', *Metacognition and Learning* 3(1), 39–58.

Bawarshi, A. S. and Mary Jo Reiff (2010), *Genre: An Introduction To History, Theory, Research, And Pedagogy*, Fort Collins, Parlor Press.

Bazeley, P. and Jackson, K. (2013), *Qualitative Data Analysis with NVivo*, London, SAGE.

Beach, R. and Friedrich, T. (2006), 'Response to writing', in C. MacArthur, S. Graham and J. Fitzgerald (eds), *Handbook of Writing Research*, New York, The Guilford Press, pp. 222–234.

Beaumont, C., Moscrop, C. and Canning, S. (2016), 'Easing the transition from school to he: scaffolding the development of self-regulated learning through a dialogic approach to feedback', *Journal of Further and Higher Education* 40(3), 331–350.

Beaumont, C., O'Doherty, M. and Shannon, L. (2011), 'Reconceptualising assessment feedback: a key to improving student learning?', *Studies in Higher Education* 36(6), 671–687.

Beauvais, C., Olive, T. and Passerault, J.-M. (2011), 'Why are some texts good and others not? relationship between text quality and management of the writing processes.', *Journal of Educational Psychology* 103(2), 415–428.

Beck, S. (2006), 'Subjectivity and intersubjectivity in the teaching and learning of writing', *Research in the Teaching of English* 40(4), 413–460.

Beck, S. W. (2009), 'Composition across secondary and post-secondary contexts: Cognitive, textual and social dimensions', *Cambridge Journal of Education* 39(3), 311–327.

Beck, S. W., Llosa, L. and Fredrick, T. (2013), 'The challenges of writing exposition: Lessons from a study of ELL and non-ELL high school students', *Reading & Writing Quarterly* 29(4), 358–380.

Behizadeh, N. and Engelhard, G. (2011), 'Historical view of the influences of measurement and writing theories on the practice of writing assessment in the united states', *Assessing Writing* 16(3), 189–211.

Bell, A., Mladenovic, R. and Price, M. (2013), 'Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars', *Assessment & Evaluation in Higher Education* 38(7), 769–788.

Benhamou, J. (2010), 'Women should mind the finals gap', *Cherwell* . URL: http://cherwell.org/2010/05/20/women-should-mind-the-finals-gap/ Accessed: 8th April 2018

Bennett, R. E. (2011), 'Formative assessment: a critical review', *Assessment in Education: Principles, Policy & Practice* 18(1), 5–25.

BERA (2011), Ethical guidelines for educational research, BERA, London.

Bereiter, C. and Scardamalia, M. (1984), 'Learning about writing from reading', *Written Communication* 1(2), 163–188.

Bereiter, C. and Scardamalia, M. (1986), 'Levels of inquiry into the nature of expertise in writing', Review of Research in Education 13(1), 259–282.

Bereiter, C. and Scardamalia, M. (1987), *The Psychology of Written Composition*, Hillsdale, NJ: Erlbaum.

Berkenkotter, C. and Huckin, T. (1995), *Genre knowledge in disciplinary communication: Cognition/culture/power*, Hillsdale, NJ: Erlbaum

Berndt, M., Strijbos, J.-W. and Fischer, F. (2018), 'Effects of written peer-feedback content and senders competence on perceptions, performance, and mindful cognitive processing', *European Journal of Psychology of Education* 33(1), 31–49.

Biggs, J. (1996), 'Enhancing teaching through constructive alignment', *Higher Education* 32(3), 347–364.

Black, P. (2015), 'Formative assessment – an optimistic but incomplete vision', *Assessment in Education: Principles, Policy & Practice* 22(1), 161–177.

Black, P. and Wiliam, D. (1998), 'Assessment and classroom learning', *Assessment in Education: Principles, Policy & Practice* 5(1), 7–74.

Black, P. and Wiliam, D. (2009), 'Developing the theory of formative assessment', *Assessment Evaluation and Accountability* 1(1), 5–31.

Bloxham, S., Boyd, P. and Orr, S. (2011), 'Mark my words: the role of assessment criteria in UK higher education grading practices', *Studies in Higher Education* 36(January), 37–41.

Bloxham, S. and West, A. (2007), 'Learning to write in higher education: students' perceptions of an intervention in developing understanding of assessment criteria', *Teaching in Higher Education* 12(1), 77–89.

Boekaerts, M. (2011), Emotions, emotion regulation, and self-regulation of learning, in D. Schunk and B. Zimmerman, (eds), *Handbook of self-regulation of Learning and Performance*, Routledge, pp. 408-425.

Boud, D. (1995), Assessment and learning: contradictory or complementary?, in P. Knight (ed.), *Assessment for Learning in Higher Education*, London, Kogan Page pp. 35–48.

Boud, D. (2000), 'Sustainable assessment: Rethinking assessment for the learning society', *Studies in Continuing Education* 22(2), 151–167.

Boud, D. (2017), Standards-based assessment for an era of increasing transparency, in D. Carless, D. Bridges, C. Chan and R. Glofcheski (eds), *Scaling up Assessment for Learning in Higher Education*, Singapore, Springer, pp. 19-31.

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

Boud, D., Lawson, R. and Thompson, D. G. (2013), 'Does student engagement in self-assessment calibrate their judgement over time?', *Assessment & Evaluation in Higher Education* 38(8), 941–956.

Boud, D. and Molloy, E. (2013), 'Rethinking models of feedback for learning: the challenge of design', *Assessment & Evaluation in Higher Education* 38(6), 698–712.

Boud, D. and Falchikov, N. (2007), 'Developing assessment for informing judgement'. In D. Boud and N. Falchikov (eds) *Rethinking Assessment in Higher Education: Learning for the Longer Term*, Abingdon, Routledge, pp. 181 - 197

Boud, D. and Soler, R. (2016), 'Sustainable assessment revisited', *Assessment & Evaluation in Higher Education*, 41 (3), 400-413.

Brinkmann, S. and Kvale, S. (2015), *InterViews*, 3rd edn, London, SAGE Publications.

Brooks, V. (2012), 'Marking as judgment', *Research Papers in Education* 27(1), 63–80.

Brown, A. (1992), 'Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings', *The Journal of the Learning Sciences* 2(2), 141–178.

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C. and Zumbrunn, S. (2013), 'Examining dimensions of self-efficacy for writing.', *Journal of Educational Psychology* 105(1), 25–38.

Bruning, R. and Kauffman, D. (2015), Self-efficacy beliefs and motivation in writing development, in C. MacArthur, S. Graham and J. Fitzgerald, (eds), *Handbook of Writing Research*, New York, Guilford Press, pp. 160–173.

Burke, D. (2009), 'Strategies for using feedback students bring to higher education', *Assessment & Evaluation in Higher Education* 34(1), 41–50.

Butler, D. and Cartier, S. (2017), Advancing research and practice about self-regulated learning: The promise of in-depth case study methodologies, in D. Schunk and J. Greene, (eds), *Handbook of self-regulation of Learning and Performance*, Abingdon, Routledge.

Butler, D. L. (2011), Investigating self-regulated learning using in-depth case studies, in B. J. Zimmerman and D. H. Schunk, (eds), *Handbook of self-regulation of Learning and Performance*, Routledge, Abingdon, pp. 346-360.

Butler, D. L. and Cartier, S. C. (2004), 'Promoting effective task interpretation as an important work habit: A key to successful teaching and learning', *Teachers College Record* 106(9), 1729–1758.

Butler, D. L. and Winne, P. H. (1995), 'Feedback and self-regulated learning: A theoretical synthesis', *Review of Educational Research* 65(3), 245–281.

Cameron, D. (2005), 'Language, gender, and sexuality: Current issues and new directions', *Applied Linguistics* 26(4), 482–502.

Camp, H. (2012), 'The psychology of writing development and its implications for assessment', *Assessing Writing* 17(2), 92–105.

Carless, D. (2009), 'Trust, distrust and their impact on assessment reform', *Assessment & Evaluation in Higher Education* 34(1), 78–89.

Carless, D. (2013a), Trust and its role in facilitating dialogic feedback. In D. Boud and E. Molloy (eds) *Feedback in Higher and Professional Education: Understanding it and Doing it Well*, Abingdon, Routledge, pp. 90-103.

Carless, D. (2013b), Sustainable feedback and the development of student self-evaluative capacities. In S. Merry, M. Price, D. Carless, and M. Taras (eds) *Reconceptualising feedback in higher education: developing dialogue with students.* London: Routledge, pp. 117-126.

Carless, D. and Chan, K. K. H. (2017), 'Managing dialogic use of exemplars', *Assessment & Evaluation in Higher Education* 42(6), 930–941.

Carless, D., Salter, D., Yang, M., and Lam, J. (2011), Developing sustainable feedback practices, *Studies in Higher Education* 36(4), 395-407.

Carroll, L. (2002), *Rehearsing New Roles: How College Students Develop as Writers*, Carbondale, Southern Illinois University Press.

Cartney, P. (2010), 'Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used', *Assessment & Evaluation in Higher Education* 35(5), 551–564.

Cassidy, S. (2007), 'Assessing 'inexperienced' students' ability to self-assess: exploring links with learning style and academic personal control', *Assessment & Evaluation in Higher Education* 32(February), 37–41.

Catt, R. and Gregory, G. (2006), The point of writing: is student writing in higher education developed or merely assessed?, in L. Ganobcsik-Williams (ed.), *Teaching Academic Writing in UK Higher Education: Theories, Practices and Models*, Basingstoke, Palgrave Macmillan, pp. 16–29.

Chambers, E. and Gregory, M. (2006), *Teaching and Learning in English Literature*, London, SAGE.

Chanquoy, L. (2016, July), Writing, spelling and emotion. EARLI SIG Writing Conference Keynote, Liverpool, UK.

Cho, K. and MacArthur, C. (2010), 'Student revision with peer and expert reviewing', *Learning and Instruction* 20(4), 328–338.

Cho, K. and MacArthur, C. (2011), 'Learning by reviewing.', *Journal of Educational Psychology* 103(1), 73–84.

Cho, K. and Schunn, C. D. (2007), 'Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system', *Computers & Education* 48(3), 409–426.

Cho, K., Schunn, C. D. and Charney, D. (2006), 'Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts', *Written Communication* 23(3), 260–294.

Cho, Y. H. and Cho, K. (2011), 'Peer reviewers learn from giving comments', *Instructional Science* 39(5), 629–643.

Christie, H., Barron, P. and D'Annunzio-Green, N. (2013), 'Direct entrants in transition: becoming independent learners', *Studies in Higher Education* 38(4), 623-637.

Christie, H., Tett, L., Cree, V. E., Hounsell, J. and McCune, V. (2008), ''a real rollercoaster of confidence and emotions': Learning to be a university student.', *Studies in Higher Education* 33(5), 567–581.

Christie, H., Tett, L., Cree, V. E. and McCune, V. (2011), 'Capital accumulation: working class students learning how to learn in HE', *Teaching in Higher Education* 16(2), 145–155.

Christie, H., Tett, L., Cree, V. E. and McCune, V. (2016), 'it all just clicked: a longitudinal perspective on transitions within university', *Studies in Higher Education* 41(3), 478–490.

Clark, I. (2012), 'Formative assessment: Assessment is for self-regulated learning', *Educational Psychology Review* 24(2), 205–249.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R. and Schauble, L. (2003), 'Design experiments in educational research', *Educational Researcher* 32(1), 9–13.

Coffey, J. E., Hammer, D., Levin, D. M. and Grant, T. (2011), 'The missing disciplinary substance of formative assessment', *Journal of Research in Science Teaching* 48(10), 1109–1136.

Court, K. (2014), 'Tutor feedback on draft essays: developing students' academic writing and subject knowledge', *Journal of Further and Higher Education* 38(3), 327–345.

Covill, A. (2012), 'College students' use of a writing rubric: Effect on quality of writing, self-efficacy, and writing practices', *Journal of Writing Assessment* 5(1).

Cresswell, M. (2000), The role of public examinations in defining and monitoring standards, in H. Goldstein and A. Heath, (eds), *Educational Standards*, Oxford, Oxford University Press, pp. 69–104.

Crossley, S., Roscoe, R. and McNamara, D. (2014), 'What is successful writing? an investigation into the multiple ways writers can write successful essays', *Written Communication* 31(2), 184–214.

Currie, R. (1994), 'The arts and social sciences, 1914-1939', in B. Harrison (ed.), *The History of the University of Oxford: Volume VIII: The Twentieth Century*, Oxford, Oxford University Press, pp. 109–138.

Darlington, E. (2014), *Changes in mathematical culture for post-compulsory mathematics students: the roles of questions & approaches to learning*. Unpublished PhD thesis, University of Oxford.

De Milliano, I., van Gelderen, A. and Sleegers, P. (2012), 'Patterns of cognitive self regulation of adolescent struggling writers', *Written Communication* 29(3), 303–325.

Deane, M. and Ganobcsik-Williams, L. (2012), 'Providing a hub for writing development: A profile of the Centre for Academic Writing (CAW), Coventry University, England', in C. Thaiss, G. Brauer, P. Carlino and L. Ganobcsik-Williams, (eds), *Writing programs worldwide: Profiles of academic writing in many places*, Fort Collins, The WAC Clearinghouse, pp. 189–201.

Devitt, A. J. (2015), 'Genre performances: John Swales' genre analysis and rhetorical linguistic genre studies', *Journal of English for Academic Purposes* 19, 44 – 51.

Dinsmore, D. L., Alexander, P. and Loughlin, S. M. (2008), 'Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning', *Educational Psychology Review* 20(4), 391–409.

Dinsmore, D. L. and Parkinson, M. M. (2013), 'What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration', *Learning and Instruction* 24, 4–14.

Dixon, H., Hawe, E. and Parr, H. (2011), 'Enacting Assessment for Learning: the beliefs practice nexus', *Assessment in Education: Principles, Policy & Practice* 18 (4), 365-379.

Dixon, Z. and Moxley, J. (2013), 'Everything is illuminated: What big data can tell us about teacher commentary', *Assessing Writing* 18(4), 241–256.

Dochy, F., Segers, M. and Sluijsmans, D. (1999), 'The use of self-, peer and co-assessment in higher education: a review', *Studies in Educational Evaluation* 24(3), 331–350.

Donahue, C. (2012), 'Transfer, portability, generalization. (How) does composition expertise "carry"?', in K. Ritter and P. Matsuda, (eds), *Exploring Composition Studies: Sites, Issues and Perspectives*, Logan, Utah State Press, pp. 145–166.

Dunlosky, J. and Thiede, K. W. (2013), 'Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement', *Learning and Instruction* 24, 58–61.

Dunning, D., Heath, C. and Suls, J. (2002), 'Flawed Self-Assessment. Implications For Health, Education And The Workplace', *Psychological Science in the Public Interest* 5(3), 69–106.

Dweck, C. S. and Master, A. (2008), 'Self-theories motivate self-regulated learning', in D. H. Schunk and B. J. Zimmerman, (eds), *Motivation and Self-Regulated Learning: Theory, Research and Applications*, New York, Routledge, pp. 31–51.

Eaglestone, R. (2007), 'Transition, 'Hard' Theory and disciplinary consciousness', *International Journal of Adolescence and Youth*, 14(1), 31-42.

Eastham, J. (2014), 'Why are male students at Oxford getting more firsts?'. *The Guardian.* URL: https://www.theguardian.com/education/mortarboard/2014/apr/24/oxford-university-firsts-gender-gap. Accessed 8th April 2018.

Efklides, A. and Volet, S. (2005), 'Emotional experiences during learning: Multiple, situated and dynamic', *Learning and Instruction* 15(5), 377–516.

Elander, J., Harrington, K., Norton, L., Robinson, H. and Reddy, P. (2006), 'Complex skills and academic writing: a review of evidence about the types of learning required to meet core assessment criteria', *Assessment & Evaluation in Higher Education* 31(1), 71–90.

Elbow, P. (1998), *Writing with power: techniques for mastering the writing process*, 2nd edition, Oxford, Oxford University Press.

Elbow, P. (2006), 'Do we need a single standard of value for institutional assessment? An essay response to Asao Inoue's community-based assessment pedagogy', *Assessing Writing* 11(2), 81–99.

Elliott, V., Baird, J., Hopfenbeck, T., Ingram., J., Thompson, I., Usher, N., Zantout, M., Richardson, J. and Coleman, R. (2016), *A marked improvement? A review of the evidence on written marking*, Education Endowment Foundation, London.

Ericsson, K. A. and Simon, H. (1993), *Protocol Analysis*, Cambridge, MA, MIT Press.

ESRC (2012), 'Framework for research ethics', ESRC, Swindon.

Evans, C. (2013), 'Making sense of assessment feedback in higher education', *Review of Educational Research* 83(1), 70–120.

Evans, K. A. (2001), 'Rethinking self-assessment as a tool for response', *Teaching English in the Two Year College* 28(3), 293–301.

Faculty of English Language and Literature, University of Oxford (2013), 'Examiners reports'. University of Oxford, Oxford.

Faculty of English Language and Literature, University of Oxford (2014), 'Examiners reports'. University of Oxford, Oxford.

Faculty of English Language and Literature, University of Oxford (2016), 'Examiners reports'. University of Oxford, Oxford.

Falchikov, N. (2005), *Improving Assessment Through Student Involvement*, Abingdon, Routledge.

Falchikov, N. and Boud, D. (2007), Assessment and emotion: the impact of being assessed, in D. Boud and N. Falchikov (eds) *Rethinking Assessment in Higher Education: Learning for the Longer Term*, Routledge, London, pp. 144–155.

Ferrari, M., Bouffard, T. and Rainville, L. (1998), 'What makes a good writer? Differences in good and poor writers' self-regulation of writing', *Instructional Science* 26(6), 473–488.

Flavell, J. H. (1979), 'Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry.', *American Psychologist* 34(10), 906–911.

Flower, L. and Hayes, J. R. (1980), 'The cognition of discovery : Defining a rhetorical problem', *College Composition and Communication* 31(1), 21–32.

Flower, L., Stein, V., Ackerman, J., Kantz, M., McCormick, K. and Peck, W. (1990), *Reading-to-Write: Exploring a Cognitive and Social Process*, New York, Oxford University Press.

Flowerdew, J. (2015), 'John Swales's approach to pedagogy in genre analysis: A perspective from 25 years on', *Journal of English for Academic Purposes* 19, 102 – 112.

Forsythe, A. and Johnson, S. (2017), 'Thanks, but no-thanks for the feedback', *Assessment & Evaluation in Higher Education* 42(6), 78–89.

Galbraith, D. (2009), 'Writing as discovery', *BJEP Monograph Series 2 (6): Teaching and Learning Writing* pp. 5–26.

Galbraith, D. and Torrance, M. (2004), Revision in the context of different drafting strategies, in L. Allal, L. Chanquoy and P. Largy, (eds), *Revision of written language: cognitive and instructional processes*, Dortrecht, Kluver, pp. 63–86.

Gardner, S. and Nesi, H. (2012), 'A classification of genre families in university student writing', *Applied Linguistics* 34(1), 25–52.

Gielen, S., Dochy, F. and Onghena, P. (2011), 'An inventory of peer assessment diversity', *Assessment & Evaluation in Higher Education* 36(2), 137–155.

Gielen, S., Dochy, F., Onghena, P., Struyven, K. and Smeets, S. (2011), 'Goals of peer assessment and their associated quality concepts', *Studies in Higher Education* 36(6), 719–735.

Gourlay, L. (2009), 'Threshold practices: becoming a student through academic literacies', *London Review of Education* 7(2), 181–192.

Gourlay, L. and Deane, J. (2012), 'Loss, responsibility, blame? staff discourses of student plagiarism', *Innovations in Education and Teaching International* 49(1), 19–29.

Graham, S. (2006), 'Writing', in P. Alexander and P. Winne, (eds), *Handbook of Educational Psychology*, Mahwah, NJ, Erlbaum, pp. 457–478.

Graham, S., Gillespie, A. and McKeown, D. (2013), 'Writing: importance, development, and instruction', *Reading and Writing* 26(1), 1–15.

Graham, S., Harris, K. and Hebert, M. (2011), *Informing writing: The benefits of formative assessment*, Carnegie Corporation of New York, New York.

Graham, S. and Harris, K. R. (2014), 'Conducting high quality writing intervention research: Twelve recommendations', *Journal of Writing Research* 6(6), 89–123.

Graham, S., Harris, K. R. and Troia, G. A. (1998), 'Writing and self-regulation: Cases from the self-regulated strategy development model', in D. H. Schunk and B. J. Zimmerman, (eds), *Self-regulated learning: From teaching to self-reflective practice*, New York, Guilford Press, pp. 20–41.

Grainger, T., Goouch, K. and Lambirth, A. (2003), 'Playing the game called writing': Children's views and voices', *English in Education* 37(2), 4–15.

Green, A. (2005), 'English literature: from sixth form to university', *International Journal of Adolescence and Youth* 12(4), 253–280.

Greene, J. and Azevedo, R. (2007), 'A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions', *Review of Educational Research* 77(3), 334–372.

Greene, J., Hutchison, L. A., Costa, L.-J. and Crompton, H. (2012), 'Investigating how college students' task definitions and plans relate to self-regulated learning processing and understanding of a complex science topic', *Contemporary Educational Psychology* 37(4), 307–320.

Greene, J., Robertson, J. and Croker Costa, L. (2011), 'Assessing self-regulated learning using think-aloud methods', in B. Zimmerman and D. H. Schunk, (eds), *Handbook of self-regulation of Learning and Performance*, Routledge, Abingdon, pp. 313–328.

Guillemin, M. and Gillam, L. (2004), 'Ethics, reflexivity, and ethically important moments in research', *Qualitative Inquiry* 10(2), 261–280.

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

Hacker, D., Keener, M. and Kircher, J. (2009), 'Writing is applied metacognition', in D. J. Hacker, J. Dunlosky and A. Graesser, (eds), *Handbook of Metacognition in Education*, New York, Routledge, pp. 154–172.

Hadwin, A. F. and Webster, E. (2013), 'Calibration in goal setting: Examining the nature of judgments of confidence', *Learning and Instruction* 24, 37–47.

Hadwin, A. F. and Winne, P. H. (2012), 'Promoting learning skills in undergraduate students', in J. Kirby and M. Lawson, (eds), *Enhancing the Quality of Learning: Dispositions, Instruction and Learning Processes*, New York, Cambridge University Press, pp. 201–227.

Hadwin, A., Jarvela, S. and Miller, M. (2011), 'Self-regulated, co-regulated and socially shared regulation of learning', in B. Zimmerman and D. H. Schunk, (eds), *Handbook of self-regulation of Learning and Performance*, Abingdon, Routledge, pp. 65–84.

Hammersley, M. (2008), 'Troubles with triangulation', in M. Bergman, (ed.), *Advances in Mixed Methods Research*, London, Sage, pp. 22–37.

Hamp-Lyons, L. (2011), 'Writing assessment: Shifting issues, new tools, enduring questions', *Assessing Writing* 16(1), 3–5.

Handley, K. and Williams, L. (2011), 'From copying to learning: using exemplars to engage students with assessment criteria and feedback', *Assessment & Evaluation in Higher Education* 36(1), 95–108.

Hanrahan, S. and Isaacs, G. (2001), 'Assessing self- and peer-assessment: The students' views', *Higher Education Research & Development* 20(1), 53–70.

Harris, K. R., Graham, S., Brindle, M. and Sandmel, K. (2009), 'Metacognition and children's writing', in D. Hacker, J. Dunlosky and A. Graesser, (eds), *Handbook of Metacognition in Education*, Abingdon, Routledge, pp. 131–153.

Hathaway, J. (2015), 'Developing that voice: locating academic writing tuition in the mainstream of higher education', *Teaching in Higher Education* 20(5), 506–517.

Hattie, J. and Timperley, H. (2007), 'The power of feedback', *Review of Educational Research* 77(1), 81–112.

Havnes, A. (2008), 'Peer-mediated learning beyond the curriculum', *Studies in Higher Education* 33(2), 193–204.

Hawe, E., Lightfoot, U. and Dixon, H. (2017), 'First-year students working with exemplars: promoting self-efficacy, self-monitoring and self-regulation', *Journal of Further and Higher Education.*

Hawthorne, K. A., Bol, L. and Pribesh, S. (2017), 'Can providing rubrics for writing tasks improve developing writers' calibration accuracy?', *The Journal of Experimental Education* 85(4), 689–708.

Hayes, J. (2004), 'What triggers revision?', in L. Allal, L. Chanquoy and P. Largy, (eds), *Revision: Cognitive and Instructional Processes*, Boston, Kluwer, pp. 9–20.

Hayes, J. and Flower, L. (1980), 'Identifying the organization of writing processes', in L. Gregg and W. Steinberg, (eds), *Cognitive Processes in Writing*, New Jersey, Erlbaum, pp. 3–30.

Hayes, J. R. (2012), 'Modeling and remodeling writing', *Written Communication* 29(3), 369–388.

Head, E. (2009), 'The ethics and implications of paying participants in qualitative research', *International Journal of Social Research Methodology* 12(4), 335–344

Hendry, G. D., Bromberger, N. and Armstrong, S. (2011), 'Constructive guidance and feedback for learning: the usefulness of exemplars, marking sheets and different types of feedback in a first year law subject', *Assessment & Evaluation in Higher Education* 36(1), 1–11.

Hidi, S. and Boscolo, P. (2006), *Writing and motivation*, Amsterdam, BRILL.

Higgins, R., Hartley, P. and Skelton, A. (2001), 'Getting the message across: The problem of communicating assessment feedback', *Teaching in Higher Education* 6(2), 269–274.

Honicke, T. and Broadbent, J. (2016), 'The influence of academic self-efficacy on academic performance: A systematic review', *Educational Research Review* 17(1), 63–84.

Horn, J. (2013), 'Signature pedagogy/powerful pedagogy: The Oxford tutorial system in the humanities.', *Arts & Humanities in Higher Education* 12(4), 350–366.

Howard, C., Barrett, A. and Frick, T. (2010), 'Anonymity to promote peer feedback: pre-service teachers' comments in asynchronous computer-mediated communication', *Journal of Educational Computing Research* 43(1), 89–112.

Hughes, G. (2011), 'Towards a personal best: a case for introducing ipsative assessment in higher education', *Studies in Higher Education* 36(February), 37–41.

Huisman, B., Saab, N., van Driel, J. and van den Broek, P. (2018), 'Peer feedback on academic writing: undergraduate students peer feedback role, peer feedback perceptions and essay performance', *Assessment & Evaluation in Higher Education*.

Hyland, K. (2003), 'Genre-based pedagogies: A social response to process', *Journal of Second Language Writing* 12(1), 17–29.

Inoue, A. B. (2004), 'Community-based assessment pedagogy', *Assessing Writing* 9(3), 208–238.

Ivanic, R. (2004), 'Discourses of writing and learning to write', *Language and Education* 18(3), 220–245.

Johnson, R. B. and Onwuegbuzie, A. J. (2004), 'Mixed methods research: A research paradigm whose time has come', *Educational Researcher* 33(7), 14–26.

Jones, S. (2011), 'Mapping the landscape: Gender and the writing classroom', *Journal of Writing Research* 3(3), 161–179.

Kaufman, J. H. and Schunn, C. D. (2011), 'Students' perceptions about peer assessment for writing: Their origin and impact on revision work', *Instructional Science* 39(3), 387–406.

Kellogg, R. (1994), *The Psychology of Writing*, New York, Oxford University Press.

Kellogg, R. (2008), 'Training writing skills: A cognitive developmental perspective', *Journal of Writing Research* 1(1), 1–26.

Kellogg, R. T. (2006), 'Professional writing expertise', in K. A. Ericsson, N. Charness, P. Feltovich and R. Hoffman, (eds), *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge, Cambridge University Press, pp. 389–402.

Kieft, M., Rijlaarsdam, G., Galbraith, D. and van den Bergh, H. (2007), 'The effects of adapting a writing course to students' writing strategies', *British Journal of Educational Psychology* 77(3), 565–578.

Kieft, M., Rijlaarsdam, G. and van den Bergh, H. (2008), 'An aptitude treatment interaction approach to writing-to-learn', *Learning and Instruction* 18(4), 379–390.

Kirby, N. F. and Downs, C. T. (2007), 'Self-assessment and the disadvantaged student: potential for encouraging self-regulated learning?', *Assessment & Evaluation in Higher Education* 32(4), 475–494.

Kitsantas, A., Winsler, A. and Huie, F. (2008), 'Self-regulation and ability predictors of academic success during college: A predictive validity study', *Journal of Advanced Academics* 20(1), 42–68.

Kitsantas, A. and Zimmerman, B. J. (2006), 'Enhancing self-regulation of practice: The influence of graphing and self-evaluative standards', *Metacognition and Learning* 1(3), 201–212.

Kollar, I. and Fischer, F. (2010), 'Peer assessment as collaborative learning: A cognitive perspective', *Learning and Instruction* 20(4), 344–348.

Kruger, J. and Dunning, D. (1999), 'Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments.', *Journal of personality and social psychology* 77(6), 1121.

Lavelle, E. (2009), 'Writing through college: self-efficacy and instruction', in R. Beard, D. Myhill, J. Riley and M. Nystrand, (eds), *The Sage Handbook of Writing Development*, London, Sage, pp. 415–421.

Lavelle, E. and Guarino, A. J. (2003), 'A multidimensional approach to understanding college writing processes', *Educational Psychology* 23(3), 295–305.

Lavelle, E. and Zuercher, N. (2001), 'The writing approaches of university students', *Higher Education* 42, 373–391.

Lea, M. (2004), 'Academic literacies: a pedagogy for course design', *Studies in Higher Education* 29(1), 37–41.

Lea, M. R. and Street, B. V. (1998), 'Student writing in higher education: an academic literacies approach', *Studies in Higher Education* 23(2), 157–172.

Lea, M. and Street, B. (2000), 'Student writing and staff feedback in higher education: an academic literacies approach', in M. Lea and B. Stierer, (eds), *Student writing in higher education: new contexts*, Milton Keynes, Open University Press, pp. 32 – 46.

Leijten, M. and Van Waes, L. (2013), 'Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes', *Written Communication* 30(3), 358–392.

Lillis, T. (1997), 'New voices in academia? The regulative nature of academic writing conventions', *Language and Education* 11(3), 182–199.

Lillis, T., Harrington, K., Lea, M. and Mitchell, S., (eds) (2015), *Working With Academic Literacies: Case Studies Towards Transformative Practice*, Fort Collins, WAC Clearinghouse.

Lillis, T. and Scott, M. (2007), 'Defining academic literacies research: Issues of epistemology, ideology and strategy', *Journal of Applied Linguistics* 4(1), 5–32.

Lillis, T. and Turner, J. (2001), 'Student writing in higher education: Contemporary confusion, traditional concerns', *Teaching in Higher Education* 6(1), 57–68.

Lindgren, E., Leijten, M. and Van Waes, L. (2011), 'Adapting to the reader during writing', *Written Language and Literacy* 14(2), 188–223.

Lindgren, E., Sullivan, K., Outakoski, H. and Westum, A. (2016, July), 'Revision at the point of inscription: an act of meaning-making or mere noise?' EARLI SIG Writing Conference Research Meeting, Liverpool.

Lines, H. (2012), 'A matter of personal taste: Teachers' constructs of writing quality in the secondary school English classroom', in C. Bazerman, C. Dean, J. Early, K. Lunsford, S. Null, P. Rogers and A. Stansell, (eds), *International Advances in Writing Research: Cultures, Places, Measures,* Fort Collins, The WAC Clearinghouse, pp. 167–187.

Lines, H. (2014), *"It's a matter of individual taste, I guess": Secondary school English teachers' and students' conceptualisations of quality in writing*, Unpublished PhD thesis, University of Exeter.

Liu, M. and Carless, D. (2006), 'Peer feedback: The learning element of peer assessment', *Teaching in Higher Education* 11(3), 279–290.

Locke, E. A. and Latham, G. P. (2002), 'Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.', *American Psychologist* 57(9), 705.

Ludemann, P. M. and McMakin, D. (2014), 'Perceived helpfulness of peer editing activities: First-year students' views and writing performance outcomes', *Psychology Learning & Teaching* 13(2), 129–136.

Luker, K. (2008), *Salsa Dancing into the Social Sciences: Research in an Age of Info-Glut*, Harvard University Press, London.

Lundstrom, K. and Baker,W. (2009), 'To give is better than to receive: The benefits of peer review to the reviewer's own writing', *Journal of Second Language Writing* 18(1), 30 –43.

MacArthur, C. and Lembo, L. (2009), 'Strategy instruction in writing for adult literacy learners', *Reading and Writing* 22(9), 1021–1039.

MacArthur, C. and Philippakos, Z. (2013), 'Self-regulated strategy instruction in developmental writing: A design research project', *Community College Review* 41(2), 176–195.

Marsden, E. and Torgerson, C. J. (2012), 'Single group, pre- and post-test research designs: Some methodological concerns', *Oxford Review of Education* 38(5), 583–616.

Marshall, B. (2017), 'The politics of testing', *English in Education* 51(1), 27–43.

Marshall, B. and Drummond, M. (2006) 'How teachers engage with Assessment for Learning: lessons from the classroom, *Assessment in Education: Principles, Policy & Practice* 21(2), 133-149.

Martin, P. (2006), *English: the condition of the subject*, Basingstoke, Palgrave Macmillan.

McCardle, L. and Hadwin, A. F. (2015), 'Using multiple , contextualized data sources to measure learners ' perceptions of their self-regulated learning', *Metacognition and Learning* 7(1), 43–75.

McConlogue, T. (2012), 'But is it fair? developing students' understanding of grading complex written work through peer assessment', *Assessment & Evaluation in Higher Education* 37(1), 113–123.

McConlogue, T. (2014), 'Making judgements: investigating the process of composing and receiving peer feedback', *Studies in Higher Education* 0(0), 1–12.

McCune, V. (2004), 'Development of first-year students' conceptions of essay writing', Higher Education 47(3), 257–282.

McGough, R. (2006), *Selected Poems*, London, Penguin.

McKenney, S. and Reeves, T. C. (2013), 'Systematic review of design-based research progress: Is a little knowledge a dangerous thing?', *Educational Researcher* 42(2), 97–100.

McNamara, D. S., Crossley, S. a. and McCarthy, P. M. (2010), 'Linguistic features of writing quality', *Written Communication* 27(1), 57–86.

Meek, S. E., Blakemore, L. and Marks, L. (2017), 'Is peer review an appropriate form of assessment in a MOOC? student participation and performance in formative peer review', *Assessment & Evaluation in Higher Education* 42(6), 1000–1013.

Mellanby, J., Zimdars, A. and Cortina-Borja, M. (2013), 'Sex differences in degree performance at the university of oxford', *Learning and Individual Differences* 26, 103–111.

Miles, M., Huberman, A. and Saldana, J. (2013), *Qualitative Data Analysis: A Methods Sourcebook*, 3rd edn, London: SAGE Publications.

Miller, P. J. (2003), 'The effect of scoring criteria specificity on peer and self-assessment', *Assessment & Evaluation in Higher Education* 28(4), 383–394.

Mitchell, S., Prior, P., Bilbro, R., Peake, K., See, B. H. and Andrews, R. (2008), 'A reflexive approach to interview data in an investigation of argument', *International Journal of Research & Method in Education* 31(3), 229–241.

Molloy, E., Borrell-Carrio, F. and Epstein, R. (2013), The impact of emotions in feedback, in D. Boud and E. Molloy, (eds), *Feedback in Higher and Professional Education: Understanding it and Doing it Well*, Abingdon, Routledge, pp. 50–71.

Moxley, J. (2013), 'Big data, learning analytics, and social assessment', *Journal of Writing Assessment* 6(1).

Moxley, J. and Eubanks, D. (2015), 'On keeping score: Instructors' vs. students' rubric ratings of 46,689 essays', *WPA: Writing Program Administration* 39, 53–80.

Myhill, D. (2005), 'Testing times: the impact of prior knowledge on written genres produced in examination settings', *Assessment in Education: Principles, Policy & Practice* 12(3), 289–300.

Myhill, D. and Jones, S. (2007), 'More than just error correction: Students' perspectives on their revision processes during writing', *Written Communication* 24(4), 323–343.

Nash, R. and Winstone, N. (2017), 'Responsibility-sharing in the giving and receiving of assessment feedback', *Frontiers in Psychology* 8(1519), 1–9.

Neal, M., O'Neill, P., Schendel, E. and Huot, B. (2007), 'Teachers' written responses to student writing. a selective annotated bibliography.', *Journal of Writing Assessment* 3(1), 61–72.

Negretti, R. (2012), 'Metacognition in student academic writing: A longitudinal study of metacognitive awareness and its relation to task perception, self-regulation, and evaluation of performance', *Written Communication* 29(2), 142–179.

Negretti, R. and Kuteeva, M. (2011), 'Fostering metacognitive genre awareness in L2 academic reading and writing: A case study of pre-service English teachers', *Journal of Second Language Writing* 20(2), 95–110.

Nelson, J. (1990), '"This was an easy assignment": examining how students interpret academic writing tasks. *Research in the Teaching of English* 24 (4), 362-396.

Nelson, M. M. and Schunn, C. D. (2009), 'The nature of feedback: How different types of peer feedback affect writing performance', *Instructional Science* 37(4), 375–401.

Nesi, H., Gardner, S., Thompson, P., Wickens, P., Forsyth, R., Heuboeck, A., Holmes, J., Hindle, D., Ebeling, S., Leedham, M. and Alsop, S. (2008), An Investigation Of Genres Of Assessed Writing In British Higher Education: Full Research Report. ESRC End Of Award Report, Res-000-23-0800, Swindon, ESRC.

Nicol, D. (2010), 'From monologue to dialogue: improving written feedback processes in mass higher education', *Assessment & Evaluation in Higher Education* 35(5), 501–517.

Nicol, D. and Macfarlane-Dick, D. (2006), 'Formative assessment and self-regulated learning : A model and seven principles of good feedback practice', *Studies in Higher Education* 31(2), 199–218.

Nicol, D. (2013) 'Resituating feedback from the reactive to the proactive,' in D. Boud and E. Molloy, (eds), *Feedback in Higher and Professional Education: Understanding it and Doing it Well*, Abingdon, Routledge, pp. 34-49.

Nicol, D., Thomson, A. and Breslin, C. (2014), 'Rethinking feedback practices in higher education: a peer review perspective', *Assessment & Evaluation in Higher Education* 39(1), 102–122.

Nielsen, K. (2014), 'Self-assessment methods in writing instruction: A conceptual framework, successful practices and essential strategies', *Journal of Research in Reading* 37(1), 1–16.

Nielsen, K. L., Hansen, G. and Stav, J. B. (2014), 'How the initial thinking period affects student argumentation during peer instruction: students' experiences versus observations', *Studies in Higher Education* 41(July), 1–15.

Odell, L., Goswami, D. and Herrington, A. (1983), The discourse-based interview: A procedure for exploring the tacit knowledge of writers in non-academic settings, in P. Mosenthal, L. Tamor and A. Walmsley, (eds), *Research on Writing: Principles and Methods*, New York, Longman, pp. 221–236.

Ofsted (2009), *English at the crossroads,* London, Ofsted. URL: http://dera.ioe.ac.uk/298/1/English%20at%20the%20crossroads.pdf . Accessed 8 April 2018.

Orsmond, P., Merry, S. and Callaghan, A. (2013), 'Communities of practice and ways to learning: charting the progress of biology undergraduates', *Studies in Higher Education* 38(6), 890–906.

Osborne, J. and Walker, P. (2014), 'Just ask teachers: Building expertise, trusting subjectivity, and valuing difference in writing assessment', *Assessing Writing* 22(1), 33–47.

OUSU (2010), *Initial Response To Oxford's Review Of Its Teaching Model*, Oxford, OUSU. URL: http://issuu.com/ousu/docs/ug teaching review report 2010. Accessed 8 April 2018.

OUSU (2013), *Education Vision. The Student View For The Future Of An Oxford Education.* Oxford, OUSU. URL: https://ousu.org/pageassets/representing-you/youreducation/EduvisFinal.pdf Accessed 8 April 2018.

OUSU (2015), *Student Written Submission*. Oxford, OUSU. URL: https://ousu.org/pageassets/representing-you/youreducation/SWS-FINALFINAL.pdf Accessed 8 April 2018.

OUSU (2016), *Report On Student Welfare, Mental Health, And Support*, Oxford, OUSU. URL: https://ousu.org/pageassets/representing-you/OUSU-Report-on-studentwelfare-mental-health-and-support.pdf Accessed 8 April 2018.

Oxford University (n.d.), 'National student survey - summary'. Oxford. URL: http://www.admin.ox.ac.uk/aad/surveys/ Accessed 22 August 2017.

Pajares, F. (2008), Motivational role of self-efficacy beliefs in self-regulated learning, in D. H. Schunk and B. J. Zimmerman, (eds), *Motivation and Self-Regulated Learning: Theory, Research and Applications*, New York, Routledge, pp. 111–139.

Pajares, F., Johnson, M. and Usher, E. (2007), 'Sources of writing self-efficacy beliefs of elementary, middle and high school students', *Research in the Teaching of English* 42(1), 104–120.

Panadero, E. (2016), Is it safe? social, interpersonal, and human effects of peer assessment: A review and future directions, in G. Brown and L. Harris, (eds), *Handbook of human and social conditions in assessment*, New York, Routledge, pp. 247–266.

Panadero, E., Alonso-Tapia, J. and Reche, E. (2013), 'Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers', *Studies in Educational Evaluation* 39(3), 125–132.

Patchan, M. M., Schunn, C. D. and Clark, R. J. (2011), 'Writing in natural sciences : Understanding the effects of different types of reviewers on the writing process', *Journal of Writing Research* 2(3), 365–393.

Patchan, M., Schunn, C. and Correnti, R. (2016), 'The nature of feedback: how peer feedback features affect students' implementation rate and quality of revisions', *Journal of Educational Psychology* 108(8), 1098–1120.

Patton, C. (2012), ''Some kind of weird, evil experiment': student perceptions of peer assessment', *Assessment & Evaluation in Higher Education* 37(6), 719–731.

Perkins, D. N. and Salomon, G. (1988), 'Teaching for transfer', *Educational Leadership* 46(1), 22–32.

Perkins, D. N. and Salomon, G. (1989), 'Are cognitive skills context-bound?', *Educational Researcher* 18(1), 16–25.

Perrenoud, P. (1998), 'From formative evaluation to a controlled regulation of learning processes. towards a wider conceptual field', *Assessment in Education: Principles, Policy & Practice* 5(1), 85–102.

Perry, N. E. (1998), 'Young children's self-regulated learning and contexts that support it.', *Journal of Educational Psychology* 90(4), 715–729.

Peterson, S. S. and Portier, C. (2014), 'Grade one peer and teacher feedback on student writing', *Education 3-13* 42(3), 237–257.

Phillips, D. (2006), Assessing the quality of design research proposals: some philosophical perspectives, in J. van den Akker, K. Gravemeijer, S. McKenney and N. Nieveen, (eds), *Educational Design Research*, Abingdon, Routledge, pp. 93–99.

Pillow, W. (2003), 'Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research', *International Journal of Qualitative Studies in Education* 16(2), 175–196.

Pintrich, P. (2000), The role of goal orientation in self-regulated learning, in M. Boekaerts, P. Pintrich and M. Zeidner, (eds), *Handbook of self-regulation*, San Diego, Academic Press, pp. 451–502.

Pitt, E. and Norton, L. (2017), 'Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it', *Assessment & Evaluation in Higher Education* 42(4), 499–516.

Plakans, L. (2010), 'Independent vs. integrated writing tasks: A comparison of task representation', *TESOL Quarterly* 44(1), 185–194.

Poitras, E. G. and Lajoie, S. P. (2013), 'A domain-specific account of self-regulated learning: The cognitive and metacognitive activities involved in learning through historical inquiry', *Metacognition and Learning* 8(3), 213–234.

Pollitt, A. (2012), 'The method of adaptive comparative judgement', *Assessment in Education: Principle, Policy & Practice* 19(3), 381–300.

Poulos, A. and Mahony, M. J. (2008), 'Effectiveness of feedback: the students' perspective', *Assessment & Evaluation in Higher Education* 33(2), 143–154.

Prat-Sala, M. and Redford, P. (2010), 'The interplay between motivation, self-efficacy, and approaches to studying', *British Journal of Educational Psychology* 80(2), 283–305.

Prat-Sala, M. and Redford, P. (2012), 'Writing essays: does self-efficacy matter? The relationship between self-efficacy in reading and in writing and undergraduate students' performance in essay writing', *Educational Psychology* 32(1), 9–20.

Price, M., Handley, K., Millar, J. and O'Donovan, B. (2010), 'Feedback: all that effort, but what is the effect?', *Assessment & Evaluation in Higher Education* 35(3), 277–289.

Pryor, J. and Crossouard, B. (2008), 'A socio-cultural theorisation of formative assessment', *Oxford Review of Education* 34(1), 1-20.

Putwain, D. W. (2009), 'Assessment and examination stress in Key Stage 4', *British Educational Research Journal* 35(3), 391–411.

Putwain, D. and Sander, P. (2014), Does the confidence of first-year undergraduate students change over time according to achievement goal profile? *Studies in Higher Education* 41(2), 381-398.

QAA (2016), 'Higher education review of the University of Oxford'. URL: http://www.qaa.ac.uk/reviews-and-reports/provider?UKPRN=10007774#.WUEbBOnTU2x Accessed 8 April 2018.

Raedts, M., Rijlaarsdam, G., Van Waes, L. and Daems, F. (2007), 'Observational learning through video-based models: impact on students writing accuracy of self-efficacy beliefs, task knowledge and writing performances', in P. Boscolo and S. Hidi, (eds), *Studies in writing, 20. Writing and motivation*, Oxford, Elsevier, pp. 219–238.

Ramaprasad, A. (1983), 'On the definition of feedback', *Behavioural Science* 28, 4-13.

Read, B., Francis, B. and Robson, J. (2001), ''Playing Safe': Undergraduate Essay Writing And The Presentation Of The Student Voice', *British Journal of Sociology of Education* 22, 387–399.

Reay, D., Crozier, G. and Clayton, J. (2009), 'Strangers In Paradise'? Working-Class Students In Elite Universities', *Sociology* 43(1), 1103–1121.

Reay, D. and Wiliam, D. (1999), ''I'll Be A Nothing': Structure, Agency And The Construction Of Identity Through Assessment', *British Educational Research Journal* 25(3), 343–354.

Reddy, Y. M. and Andrade, H. (2010), 'A review of rubric use in higher education', *Assessment & Evaluation in Higher Education* 35(4), 435–448.

Reiff, M. and Bawarshi, A. (2011), 'Tracing discursive resources: How students use prior genre knowledge to negotiate new writing contexts in first-year composition', *Written Communication* 28(3), 312–337.

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Raedts, M., Steendam, E. V., Toorenaar, A., and Van Den Bergh, H. (2008), 'Observation of peers in learning to write practise and research', *Journal of Writing Research* 1(1), 53–83.

Rijlaarsdam, G., Couzijn, M. and Van den Bergh, H. (2004), The study of revision as a writing process and as a learning-to-write process, in L. Allal, L. Chanquoy and P. Largy, (eds), *Revision: Cognitive and Instructional Processes*, Amsterdam, Kluwer, pp. 189–207.

Robson, C. (2011), *Real world research: a resource of users of social research methods in applied settings*, Chichester, Wiley.

Robson, S. (2016), 'Self-regulation and metacognition in young children: Does it matter if adults are present or not?', *British Educational Research Journal* 42(2), 185–206.

Rotsaert, T., Panadero, E., and Schellens, T. (2018), Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills, *European Journal of Psychology of Education* 33, 75-99.

Russell, D. R., Lea, M., Parker, K., Street, B. and Donahue, C. (2009), Exploring notions of genre in academic literacies and writing across the curriculum: Approaches across countries and contexts, in C. Bazerman, A. Bonini and D. Figueiredo, (eds), *Genre in a Changing World*, Fort Collins, Colorado, The WAC Clearinghouse, pp. 395–408.

Rust, C., Price, M. and O'Donovan, B. (2003), 'Improving students' learning by developing their understanding of assessment criteria and processes', *Assessment & Evaluation in Higher Education* 28(2), 147–164.

Ryan, T. and Henderson, M. (2017), 'Feeling feedback: students' emotional responses to educator feedback', *Assessment & Evaluation in Higher Education*.

Sadler, D. R. (1985), 'The origins and functions of evaluative criteria', Educational Theory 35(3), 285–297.

Sadler, D. R. (1987), 'Specifying and promulgating achievement standards', *Oxford Review of Education* 13(2), 191–209.

Sadler, D. R. (1989), 'Formative assessment and the design of instructional systems', *Instructional Science* 18(2), 119–144.

Sadler, D.R. (2002), Ah!... so that's 'quality', in P. Schwartz and G. Webb, (eds), *Assessment: Case Studies, Experience And Practice From Higher Education*, London, Kogan Page, pp. 130–136.

Sadler, D.R. (2007), 'Perils in the meticulous specification of goals and assessment criteria', *Assessment in Education, Principles, Policy & Practice* 14(3), 387-392.

Sadler, D. R. (2009a), 'Transforming holistic assessment and grading into a vehicle for complex learning' in G. Joughin (eds) *Assessment, Learning and Judgement in Higher Education,* pp. 45-63, Dordrecht, Springer.

Sadler, D. R. (2009b), 'Indeterminacy in the use of pre-set criteria for assessment and grading', *Assessment & Evaluation in Higher Education* 34(2), 159–179.

Sadler, D. R. (2010), 'Beyond feedback: developing student capability in complex appraisal', *Assessment & Evaluation in Higher Education* 35(5), 535–550.

Sadler, D.R. (2013),Opening up feedback, in S.Merry, M.Price, D.Carless and M. Taras (eds), *Reconceptualising feedback in higher education: Developing dialogue with students*, Abingdon, Routledge, pp. 54–63.

Sadler, D.R. (2014), 'The futility of attempting to codify academic achievement standards', *Higher Education* 67(3), 273–288.

Schunk, D. H. (2003), 'Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation', *Reading & Writing Quarterly* 19(2), 159–172.

Schunk, D. H. and Zimmerman, B. J. (1997), 'Social origins of self-regulatory competence', *Educational Psychologist* 32(4), 195–208.

Schunk, D. and Usher, E. (2011), Assessing self-efficacy for self-regulated learning, in B. Zimmerman and D. Schunk, (eds), *Handbook of self-regulation of Learning and Performance*, Abingdon, Routledge, pp. 282–297.

Schunn, C. (2016), 'Writing to learn and learning to write through sword', in S.A. Crossley & D.S. McNamara (Eds.), *Adaptive Educational Technologies for Literacy Instruction,* pp. 243–59, . NY: Taylor & Francis, Routledge

Schunn, C., Godley, A. and Demartino, S. (2016), 'The reliability and validity of peer review of writing in high school', *Journal of Adolescent & Adult Literacy* 60(1), 1–11.

Schwartz, D., Chase, C. and Bransford, J. (2012), 'Resisting overzealous transfer: Coordinating previously successful routines with needs for new learning', *Educational Psychologist* 47(3), 204–214.

Scoles, J., Huxham, M. and McArthur, J. (2013), 'No longer exempt from good practice: using exemplars to close the feedback gap for exams', *Assessment & Evaluation in Higher Education* 38(6), 631–645.

Shute, V. J. (2008), 'Focus on formative feedback', *Review of Educational Research* 78(1), 153–189.

Slavin, R. E. (2004), 'Education research can and must address "what works" questions', *Educational Researcher* 33(1), 27–28.

Slomp, D. H. (2012), 'Challenges in assessing the development of writing ability: Theories, constructs and methods', *Assessing Writing* 17(2), 81–91.

Slomp, D. H. and Fuite, J. (2004), 'Following Phaedrus: Alternate choices in surmounting the reliability/validity dilemma', *Assessing Writing* 9(3), 190–207.

Smith, K. and Hopkins, C. (2005), 'Great expectations: sixth-formers' perceptions of teaching and learning in degree-level English', *Arts and Humanities in Higher Education* 4(3), 304–318.

Snapper, G. (2007), 'Beyond Curriculum 2000: Some national and international perspectives on a level English literature', *International Journal of Adolescence and Youth* 14(1), 13–29.

Soderstrom, N. and Bjork, R. (2015), 'Learning versus performance: An integrative review', *Perspectives on Psychological Science* 10(2), 176–199.

Sommers, N. and Saltz, L. (2004), 'The novice as expert : the freshman year writing', *College Composition and Communication* 56(1), 124–149.

St. Pierre, E. A. and Jackson, A. Y. (2014), 'Qualitative data analysis after coding', *Qualitative Inquiry* 20(6), 715–719.

Stobart, G. (2012), Validity in formative assessment, in J. Gardner, ed., *Assessment and Learning*, 2nd edn, London, SAGE Publications, pp. 233–242.

Strijbos, J.-W., Narciss, S. and D¨unnebier, K. (2010), 'Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?', *Learning and Instruction* 20(4), 291–303.

Swales, J. M. (1990), *Genre Analysis*, Cambridge, Cambridge University Press.

Sword, H. (2012), *Stylish Academic Writing*, Cambridge, MA, Harvard University Press.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., and Panadero, E. (2017), 'Developing evaluative judgement: enabling students to make decisions about the quality of work'. *Higher Education.*

Tapp, J. (2013), 'I Actually Listened, I'm Proud Of Myself. The Effects Of A Participatory Pedagogy On Students' Constructions Of Academic Identities', *Teaching in Higher Education* 19(4), 323–335.

To, J. and Carless, D. (2015), 'Making productive use of exemplars: Peer discussion and teacher guidance for positive transfer of strategies', *Journal of Further and Higher Education* 40(3), 1–19.

Topping, K. (1998), 'Peer assessment between students in colleges and universities', *Review of Educational Research* 68(3), 249–276.

Torrance, H. (2012), 'Formative assessment at the crossroads: conformative, deformative and transformative assessment', *Oxford Review of Education* 38(3), 323–342.

Torrance, H. (2018), 'The return to final paper examining in English national curriculum assessment and school examinations: Issues of validity, accountability and politics', *British Journal of Educational Studies* 66(1), 3–27.

Torrance, H. and Pryor, J. (1998), *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*, Buckingham, Open University Press.

Torrance, M., Thomas, G. V. and Robinson, E. (2000), 'Individual differences in undergraduate essay-writing strategies : A longitudinal study', *Higher Education* 39, 181–200.

Unistats (n.d.), 'Ba (Hons) English Language And Literature'. URL: http://unistats.ac.uk/subjects/satisfaction/10007774FT-Q300. Accessed 20 August 2017.

Usher, N. (2013), *Assessment for learning in secondary writing classrooms*, Unpublished Master's thesis, University of Oxford.

Valsiner, J. (2000), 'Data as representations: contextualizing qualitative and quantitative research strategies', *Social Science Information* 39(1), 99–113.

van den Berg, I., Admiraal, W. and Pilot, A. (2006), 'Peer assessment in university teaching: evaluating seven course designs', *Assessment & Evaluation in Higher Education* 31(1), 19–36.

van den Bergh, H. and Rijlaarsdam, G. (2001), 'Changes in cognitive activities during the writing process and relationships with text quality', *Educational Psychology* 21(4), 373–385.

van Dinther, M., Dochy, F. and Segers, M. (2011), 'Factors affecting students' self-efficacy in higher education', *Educational Research Review* 6(2), 95–108.

van Zundert, M., Sluijsmans, D. and van Merrienboer, J. (2010), 'Effective peer assessment processes: Research findings and future directions', *Learning and Instruction* 20(4), 270–279.

Varner, L., Roscoe, R. and McNamara, D. (2013), 'Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis', *Journal of Writing Research* 5(1), 35–59.

Vermunt, J. D. and Verloop, N. (1999), 'Congruence and friction between learning and teaching', *Learning and Instruction* 9(3), 257–280.

Walker, M. (2015), 'The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback', *Assessment & Evaluation in Higher Education* 40(2), 232–247.

Wallace, D. and Hayes, J. (1991), 'Redefining revision for freshmen', *Research in the Teaching of English* 25(1), 54–66.

Wallace, D. L., Hayes, J. R., Hatch, J. a., Miller, W., Moser, G. and Silk, C. M. (1996), 'Better revision in eight minutes? Prompting first-year college writers to revise globally.', *Journal of Educational Psychology* 88(4), 682–688.

Wardle, E. (2007), 'Understanding transfer from FYC : Preliminary results of a longitudinal study', *WPA: Writing Program Administration* 31 (1-2), 65–85.

Warikoo, N. K. and Fuhr, C. (2014), 'Legitimating status: Perceptions of meritocracy and inequality among undergraduates at an elite British university', *British Educational Research Journal* 40(4), 699–717.

Weaver, M. R. (2006), 'Do students value feedback? Student perceptions of tutors' written responses', *Assessment & Evaluation in Higher Education* 31(3), 379–394.

Webster, E. and Hadwin, A. (2014), 'Emotions and emotion regulation in undergraduate studying : examining students ' reports from a self-regulated learning perspective', *Educational Psychology* 35(7), 794–818.

Wilder, L. and Wolfe, J. (2009), 'Sharing the tacit rhetorical knowledge of the literary scholar: The effects of making disciplinary conventions explicit in undergraduate writing about literature courses', *Research in the Teaching of English* 44(2), 170–209.

Wiliam, D. (2011), 'What is assessment for learning?', *Studies in Educational Evaluation* 37(1), 3–14.

Williams, Y. and Williams, D. (2017), 'How accurate can A-level English literature marking be?', *English in Education* 51(3), 255–274.

Wilson, F., Child, S. and Suto, I. (2016), 'Assessing the transition between school and university: Differences in assessment between A level and university in English', *Arts and Humanities in Higher Education* 16(2), 188–208.

Wimshurst, K. and Manning, M. (2013), 'Feed-forward assessment, exemplars and peer marking: evidence of efficacy', *Assessment & Evaluation in Higher Education* 38(4), 451–465.

Wingate, U. (2006), 'Doing away with study skills", *Teaching in Higher Education* 11(4), 457–469.

Wingate, U. (2010), 'The impact of formative feedback on the development of academic writing', *Assessment & Evaluation in Higher Education* 35(5), 519–533.

Wingate, U. (2012), ''Argument!' helping students understand what essay writing is about', *Journal of English for Academic Purposes* 11(2), 145–154.

Wingate, U. and Tribble, C. (2012), 'The best of both worlds? Towards an English for academic purposes/academic literacies writing pedagogy', *Studies in Higher Education* 37(4), 481–495.

Winne, P.H. (2005) 'A perspective on state-of-the-art research on self-regulated learning'. *Instructional Science* 33, 559-565.

Winne, P. H. (2014), 'Issues in researching self-regulated learning as patterns of events', *Metacognition and Learning* 9(2), 229–237.

Winne, P. H. and Hadwin, A. F. (1998), 'Studying as self-regulated learning', in D. J. Hacker, J. Dunlosky and A. Graesser, (eds), *Metacognition in Educational Theory and Practice*, Mahwah, NJ, Erlbaum, pp. 277–304.

Winne, P. H. and Hadwin, A. F. (2008), 'The weave of motivation and self-regulated learning', in D. H. Schunk and B. J. Zimmerman, (eds), *Motivation and Self-regulated Learning: Theory, Research and Applications,* New York, Routledge, pp. 297–315.

Winne, P. and Muis, K. (2011), 'Statistical estimates of learners' judgments about knowledge in calibration of achievement', *Metacognition and Learning* 6(2), 179–193.

Winne, P. and Perry, N. (2000), 'Measuring self-regulated learning', in M. Boekaerts, P. Pintrich and M. Zeider, (eds), *Handbook of self-regulation*, San Diego, Academic Press, pp. 531–566.

Winstone, N., Nash, R., Parker, M. and Rowntree, J. (2017), 'Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes', *Educational Psychologist* 52(1), 17–37.

Winstone, N., Nash, R., Rowntree, J. and Parker, M. (2017), 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience, *Studies in Higher Education*, 42(11), 2026-2041.

Woods, M., Macklin, R. and Lewis, G. K. (2016), 'Researcher reflexivity: exploring the impacts of CAQDAS use', *International Journal of Social Research Methodology* 19(4), 385–403.

Woolf, V. (1929), *The Letters Of Virginia Woolf: A Reflection Of The Other Person: 1929-1931*, London, Hogarth Press.

Wyatt-Smith, C. and Klenowski, V. (2013), 'Explicit, latent and meta-criteria: types of criteria at play in professional judgement practice', *Assessment in Education: Principles, Policy & Practice* 20(1), 35–52.

Yancey, K. (1999), 'Looking back as we look forward: Historicizing writing assessment', *College Composition and Communication* 50(3), 483–503.

Yin, R. (2014), *Case Study Research: Design and Methods, Applied Social Research Methods*, London, SAGE Publications.

Yorke, M. (2003), 'Formative assessment in higher education : Moves towards theory and the enhancement of pedagogic practice', *Higher Education* 45(4), 477–501.

Yu, S. and Lee, I. (2016), 'Peer feedback in second language writing (2005-2014)', *Language Teaching* 49(4), 461493.

Zimmerman, B. (2008), 'Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects', *American Educational Research Journal* 45(1), 166–183.

Zimmerman, B. (2011), 'Motivational sources and outcomes of self-regulated learning and performance', in B. Zimmerman and D. Schunk, (eds), *Handbook of self-regulation of Learning and Performance*, Abingdon, Routledge, pp. 49–64.

Zimmerman, B. J. (2000), 'Attaining self-regulation: a social cognitive perspective', in M. Boekaerts, P. R. Pintrich and M. Zeider, (eds), *Handbook of self-regulation*, San Diego, Academic Press, pp. 13–39.

Zimmerman, B. J. and Kitsantas, A. (2002), ‘Acquiring writing revision and self-regulatory skill through observation and emulation.’, *Journal of Educational Psychology* 94(4), 660–668.

Zimmerman, B. J. and Kitsantas, A. (2006), ‘A writer’s discipline: the development of self-regulatory skill’, in S. Hidi and P. Boscolo, (eds), *Writing and Motivation*, Amsterdam, Kluwer, pp. 51–69.

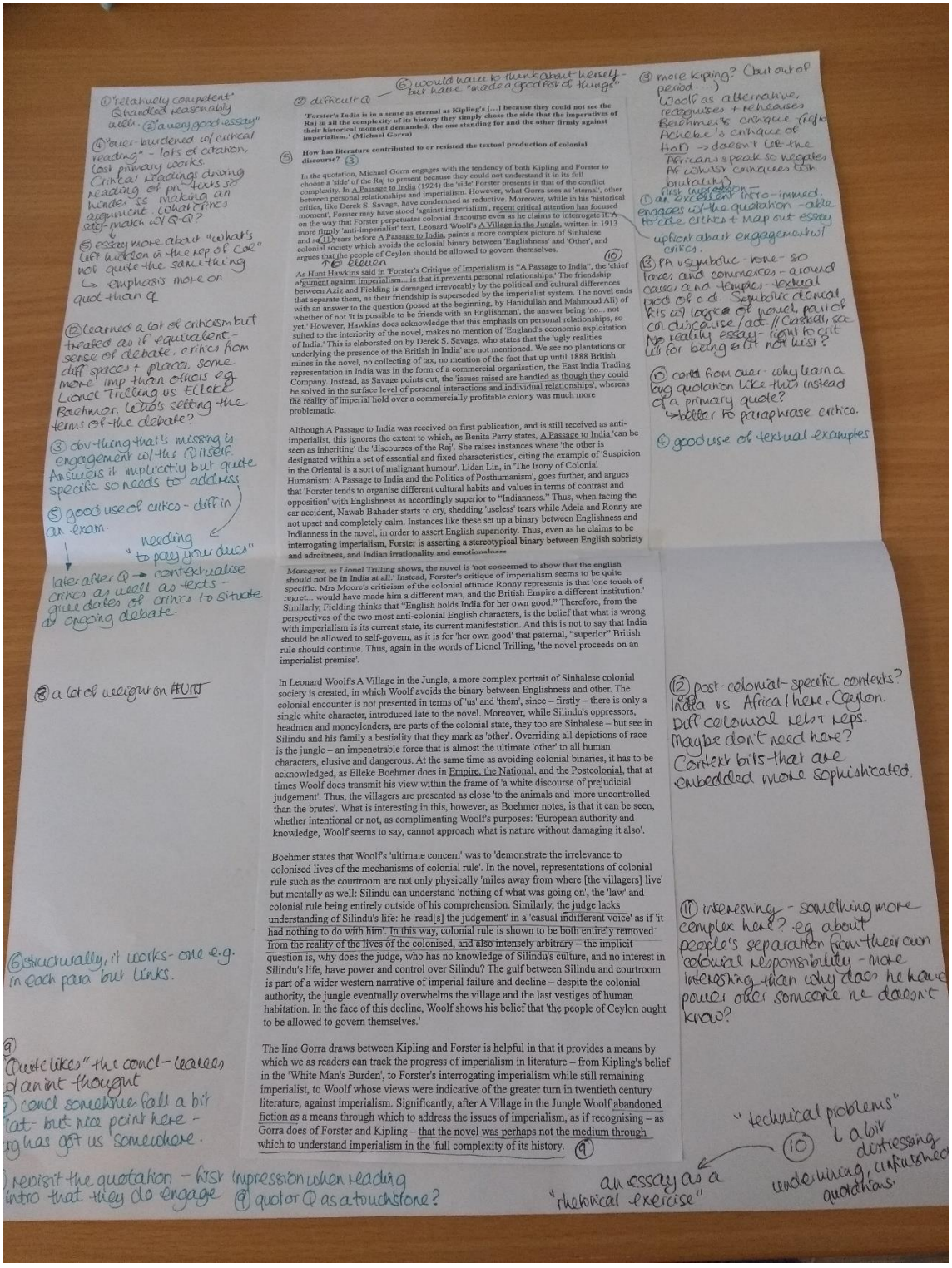
Zimmerman, B. and Kitsantas, A. (1999), ‘Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals’, *Journal of Educational Psychology* 91(2), 241–250.

Zimmerman, B. and Risemberg, R. (1997), ‘Becoming a self-regulated writer: a social cognitive perspective’, *Contemporary Educational Psychology* 22(1), 73–101.

APPENDICES

Appendix A: Workshop materials

Design process: annotating an example essay with tutor interviews (colour-coded by tutor)



Sample teaching notes

Essay 2: Forster and Woolf

Strong introduction

- Engages with the quotation and refers to the question
- Maps out essay

Conclusion

- Returns to the quotation in the conclusion
- Opens things up – we feel like we've got somewhere

Use of the quotation and question

- Interpretation of the quotation
- Question not really addressed apart from in intro
- But difficult Q – tension between specific quotation and more general Q

Use of critics

- Contextualising (e.g.) with dates
- Debates between critics
- Who's important?
- Balance between critics / textual detail / student's voice
- In some places, the writer doesn't respond to or build on critics (e.g. after Parry in Paragraph 3)
- Critics in topic sentences
- No need to quote critics as such length

Structure

- Works – deals with novels in separate paragraphs, but there is a link / transition and sense of progression

Referencing / conventions – needs to be addressed for at least one of the essays we discuss

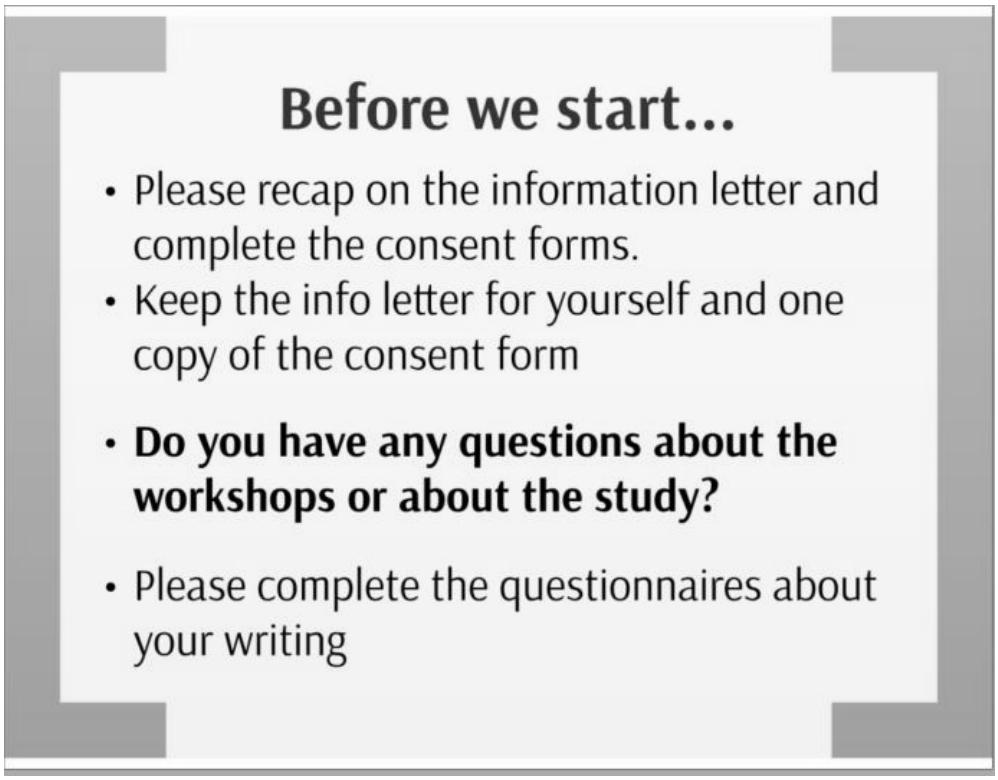
- No need to cite titles of criticism
- No need for long critical quotations
- Consistent underlining of novel titles

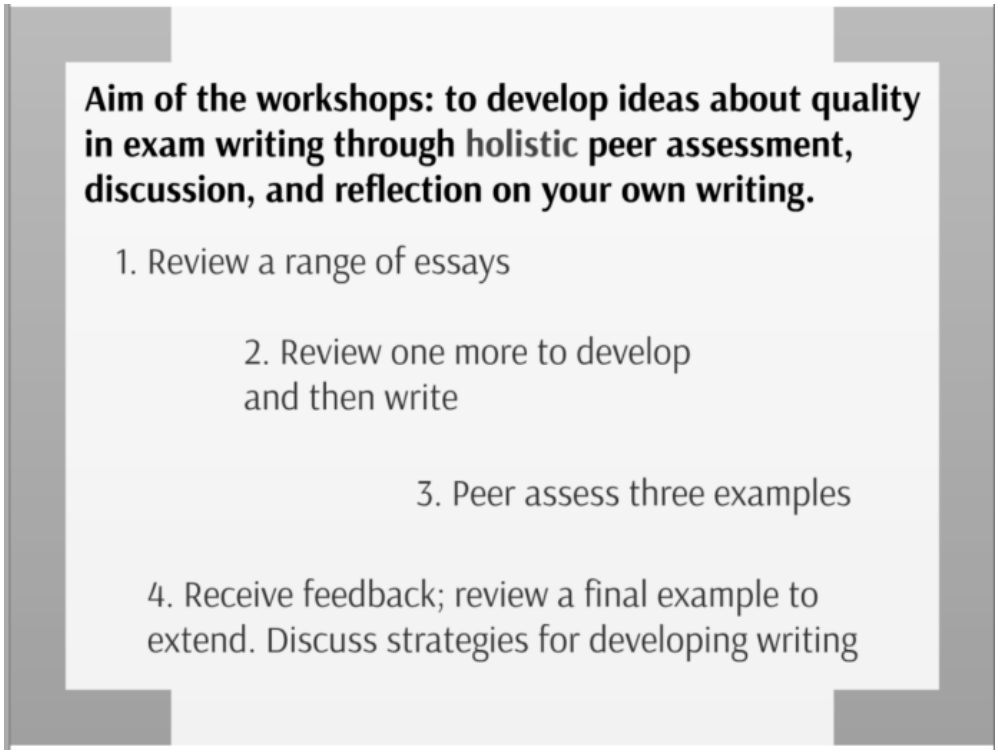
Length

- Comparatively long essay – 1230 words whereas the mean word count for essays is about 900 words (both 1st year sample and past Prelims essays sample).

Workshop slides & reflections handouts

Workshop 1





Scope of the workshops

We'll look at a range of essays all the way up to low-firsts

But... we won't discuss what mark or classification each could have got. This is a whole other skill that assessors develop after they have learned how to assess holistically or with criteria.

In the final session, we'll discuss strategies for developing your writing - *how* to put things into practice

All of the essays were written for collections or practice by students now in their third year or beyond. I've discussed the essays with four English tutors.

Criteria-first assessment

- Taking a pre-defined, fixed list of criteria, and checking whether each is absent or present, or making a statement on each.
- Can take the form of a checklist
- The text is effectively broken down into little bits when you read it

But:

- For complex pieces of work, like essays, no list of criteria is exhaustive so you can miss things that are important but not captured in the criteria
- It can also be hard to separate out criteria
- Also, if the criteria aren't clear, how do you apply them?

A **top-down** approach - see how the essay fits a set of pre-determined criteria.

Holistic assessment

- Which aspects of the essay stand out and contribute to the quality of the essay **as a whole?** Either because they really contribute to the quality of the essay, or because they hold it back?
- Deeper reading - gives scope for considering how aspects of the whole go together.
- For complex and creative pieces of work, allows for us to notice and consider unusual aspects that haven't been pre-empted, and not get caught up in paying lots of attention to criteria that don't really matter to this work.

A **bottom-up** approach - criteria emerge from reading the essay and are not pre-defined.

Today's workshop: Developing ideas about quality through reviewing a range of essays

Read the essay and respond.
Write comments of your own, explaining and justifying what you think

→ Discuss the essay as a group

Once we've looked at three, we'll discuss the key criteria that have emerged

→ Reflect on your own writing

Practicalities

When you're writing your initial response, please use **black** pen
When you make notes on our discussion, please use a **coloured** pen

The handouts are on carbonless paper - try and keep the copies together. At the end, please hand in the top copy to me and keep the yellow one for yourself. If the quality of the yellow copy isn't great, I can photocopy it for you and bring it next time (or scan it / pidge it).

Read the essay and write a response. Try to focus on ***how*** the student writes about literature rather than ***what*** the student writes.

Questions you ***could*** think about when you're writing your response to the essay:

- What stands out as being particularly successful about this essay?
- What, if anything, do you find confusing?
- What do you like about it?
- What stands out as being less developed?
- Which parts are particularly interesting or convincing?
- What's holding this essay back?

What's a response?

- Write about what you notice about the essay
- Write in full sentences and try to explain what you mean
- Feel free to annotate
- It's a response, so you can of course use 'I'
- You can also write about what you liked, what you found confusing, hard to follow...
- It doesn't need to be polished. And it's not feedback, so you don't need to think too much about the original writer, or give suggestions (although if any come to mind, that's absolutely fine)

Bringing it all together

Which aspects of writing have been important in our discussion of the example essays?

If we were to treat these aspects of writing as criteria, what would we call them?

Workshop 1 Reflections

Workshop 1: Reflecting on your own writing

1. In your opinion, which was the strongest piece of writing we read today? And which was the weakest?

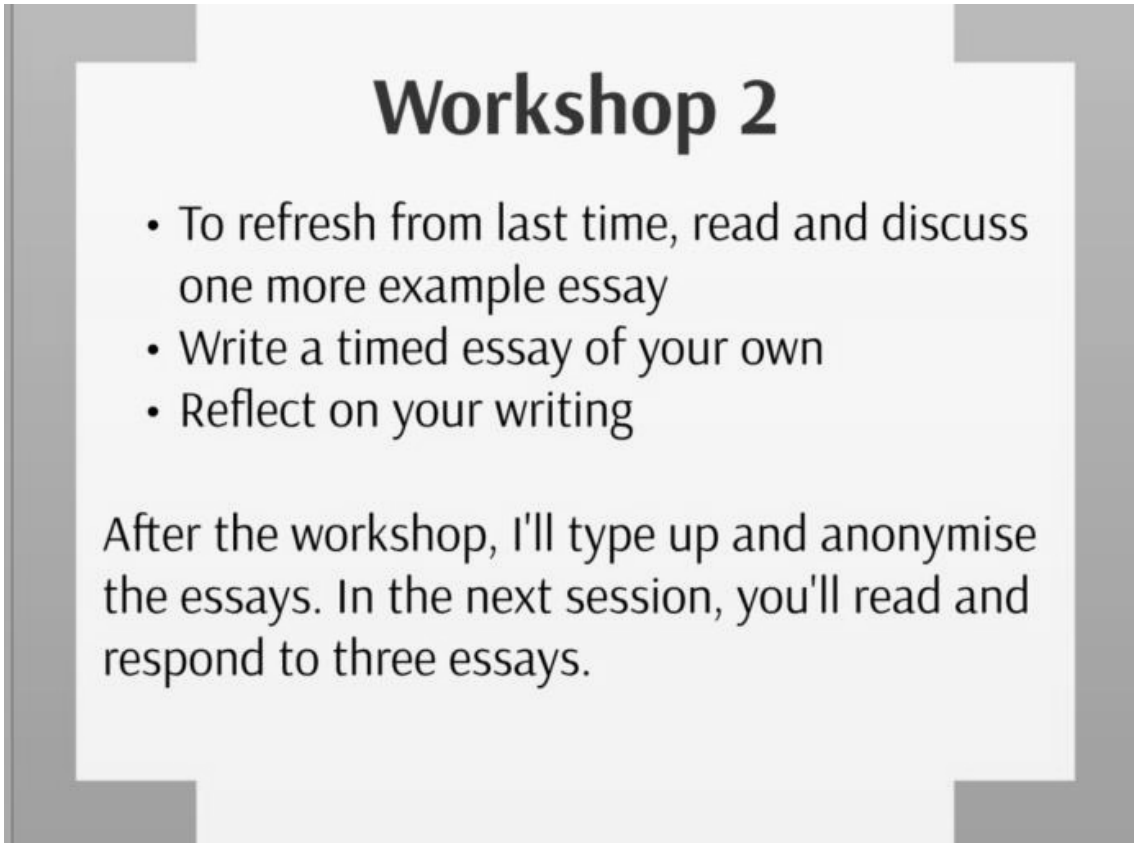
2. Did anything surprise you about the examples that we reviewed?

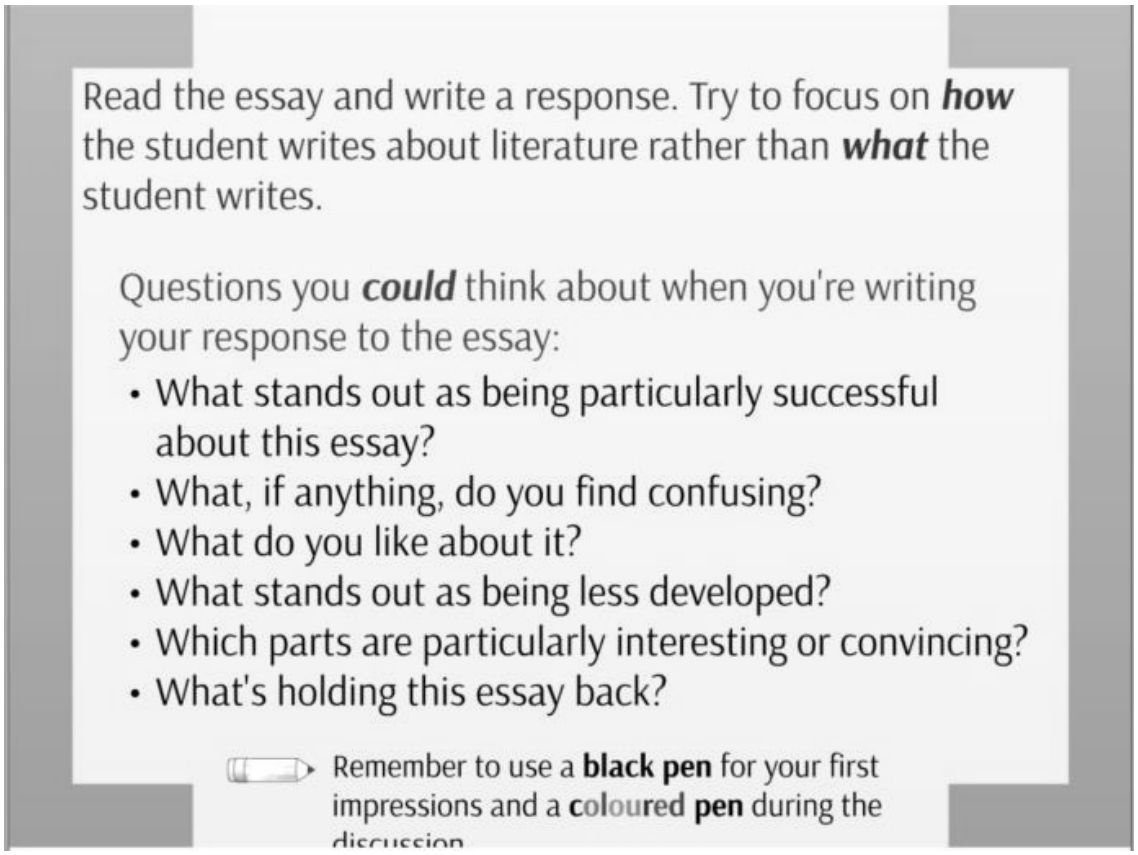
3. Did anything surprise you about the criteria that we discussed?

4. What do you feel were the most important criteria that came up in today's workshop?

5. Do you recognise any of the strengths and limitations of the example essays in your own writing? Try to point to specific essays you've written.

Workshop 2


Workshop 2
• To refresh from last time, read and discuss one more example essay
• Write a timed essay of your own
• Reflect on your writing
After the workshop, I'll type up and anonymise the essays. In the next session, you'll read and respond to three essays.


Read the essay and write a response. Try to focus on how the student writes about literature rather than what the student writes.
Questions you could think about when you're writing your response to the essay:
• What stands out as being particularly successful about this essay?
• What, if anything, do you find confusing?
• What do you like about it?
• What stands out as being less developed?
• Which parts are particularly interesting or convincing?
• What's holding this essay back?
Remember to use a black pen for your first impressions and a coloured pen during the

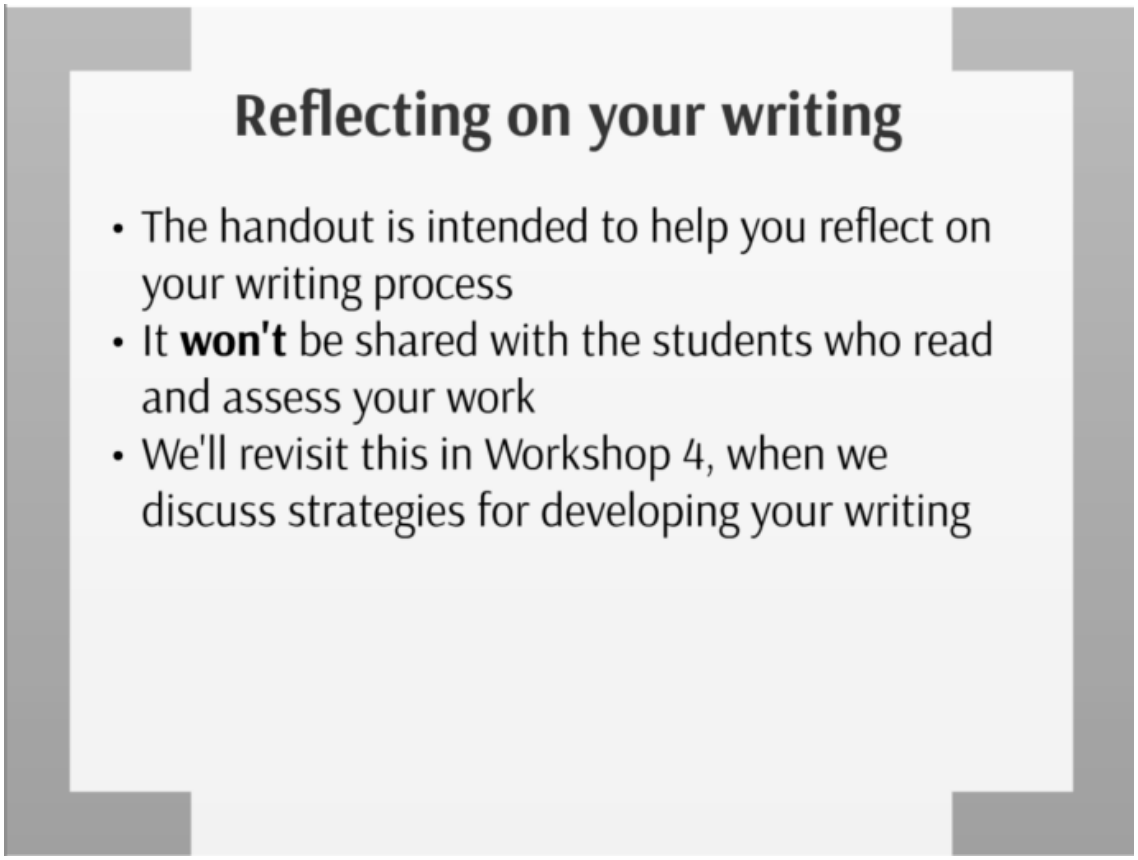
Refreshing the criteria

Thinking about the criteria you generated last time...

Would you like to add to these or clarify any of these on the basis of your review of this essay?

Writing: practicalities

- If you're doing Paper 3, older past papers refer to 'Victorian literature' and 'Victorian writers' etc. Feel free to think in terms of 'writing of the period' or 'writers of the period' instead.
- I'll give you a copy of your essay next session.



Workshop 2 Reflective Questions

(Completed immediately post-writing)

1. How did you come to choose the question you wrote on?

2. Briefly describe what you did if / when you planned.

3. How did you do on timing? Did you get the opportunity to read over your work and / or revise anything at the end?

4. What was the question (and quotation) asking you to do?

5. As you wrote, were there any changes or shifts in the way that you thought about the question and quotation?

If so, how?

6. Were you able to use any points or arguments you've written about before in essays? If so, please give some brief details.

7. Did any of these points or arguments *change or shift* as you thought about the question? If so, please give some brief details.

8. Did you think of any *new* points or arguments today as you wrote? If so, please give some brief details.

9. What do you like most about this essay? What do you feel are its strengths?

10. What do you feel are its limitations or areas for development?

11. If you had 5-10 minutes available at the end of the exam to go back to this essay, what would you do with that time?

12. Did you try anything new today? If so, what? And how did it work out?

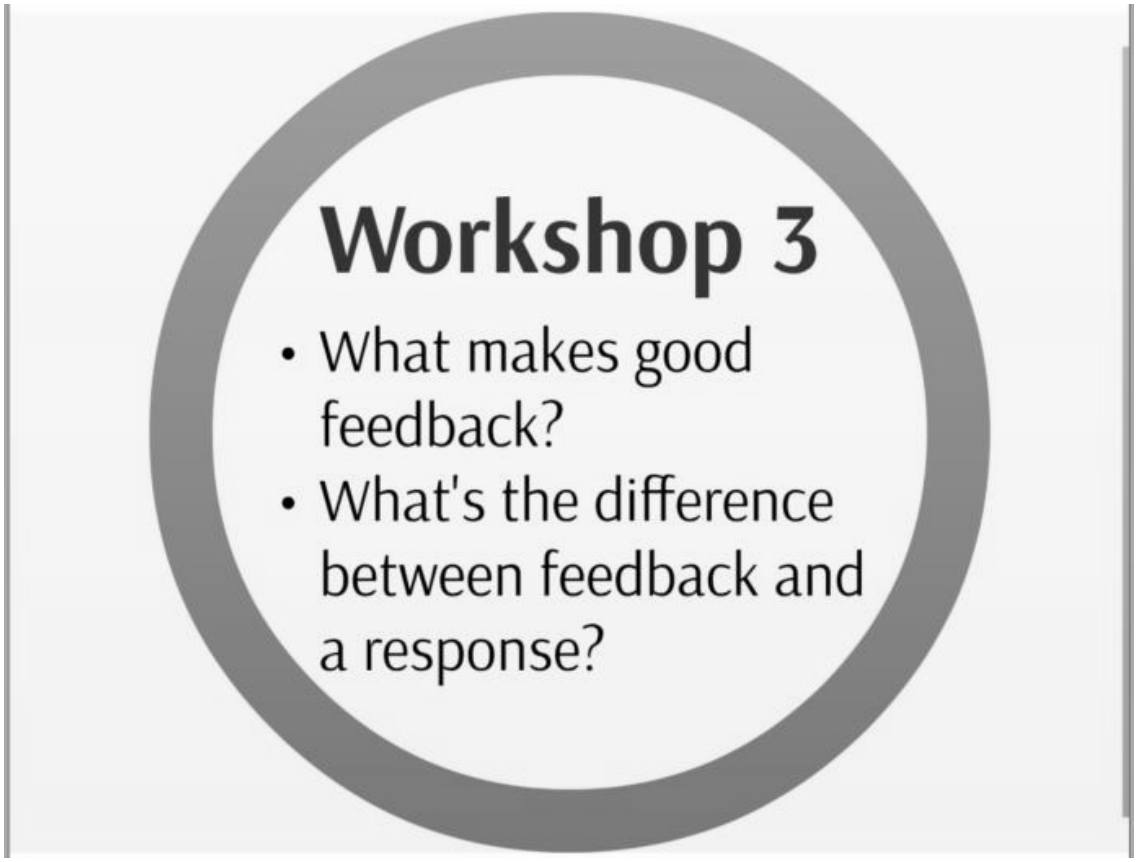
13. Thinking about the approach you took to today's essay, would you change anything about it next time?

14. As you were writing, were you thinking of a particular audience or reader?

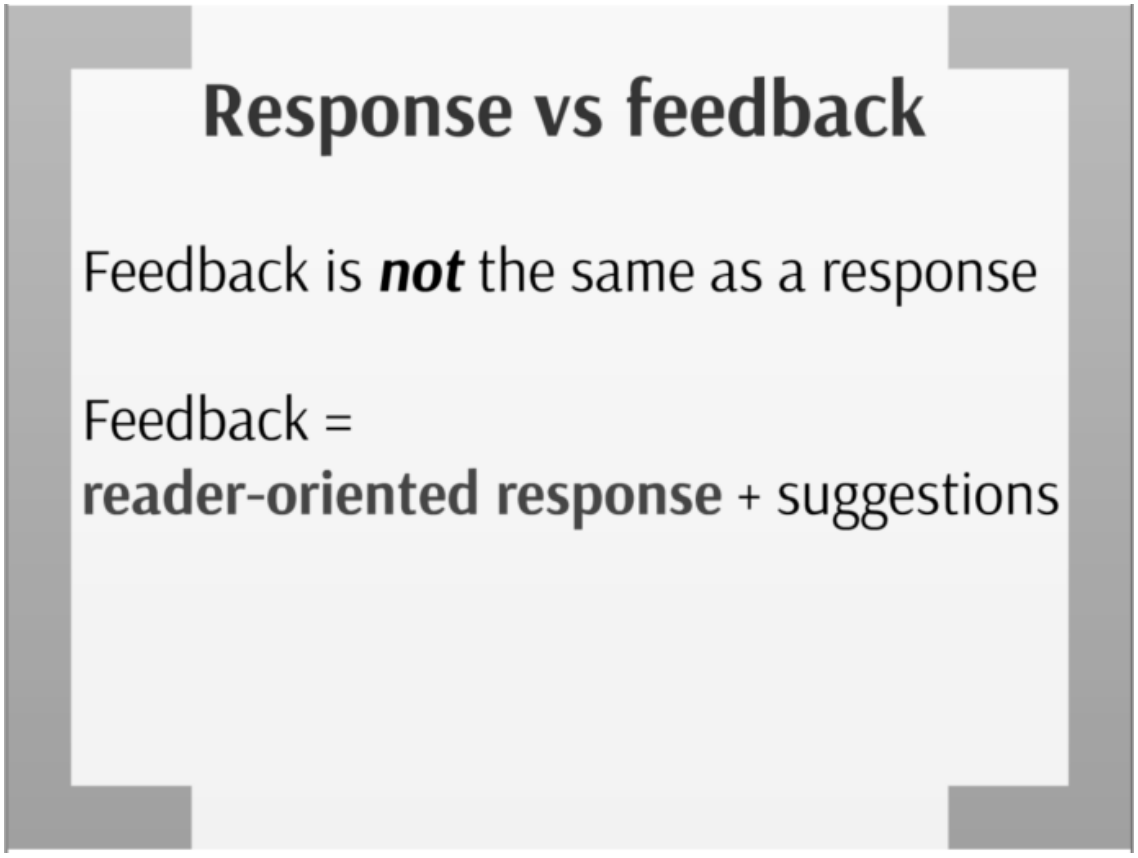
15. Did you do any revision or preparation for today's essay? If so, please give brief details (e.g. topic / authors / texts, revision techniques).

16. Anything else that comes to mind about the process of writing today's essay?

Workshop 3



Response or feedback?
5/20 dire


Response vs feedback
Feedback is *not* the same as a response
Feedback =
reader-oriented response + suggestions

'5/20 dire'
deficit perspective

developmental
perspective

Turning your response into feedback

With the example essays, we've been really thorough, pulling out *everything* that could possibly be improved or changed - this was to raise your critical awareness

With feedback, it's about channelling your critical awareness to be helpful to the writer. Pointing out everything could be really overwhelming and unrealistic.

It's important to pick out the most important things - remember that the holistic approach is about first noticing what stands out to you and then thinking about criteria.

Ground rules

- Respond as a reader – if you found a section really interesting, then say. Equally, if a section confuses you, then say.
- It's absolutely fine to use 'I' and 'like'
- Consider the tone of your feedback – try to be constructive and helpful.
- Try to explain what you mean and/or give examples
- Make suggestions in a constructive and tactful way. Keep suggestions to the most important things that could really lift the writing.
- Keep the focus on *how* the essay is written, not the content of the essay

Practicalities

- Feel free to annotate and write notes on the essays
- I've provided a separate handout for feedback, and there's more space for notes on the back
- If you want to share one of your annotations with the reader, or highlight something, please mark it up in coloured pen or highlighter
- You don't get to keep the essays, but I'm happy to photocopy your comments and notes, if you'd like.
- All of the example essays are obviously yours to keep, but please don't share them with others beyond this group.

Workshop 3: Reflective questions

(Completed after giving feedback)

1. Overall, what do you think about today's essays? Did anything surprise you about the essays you read?

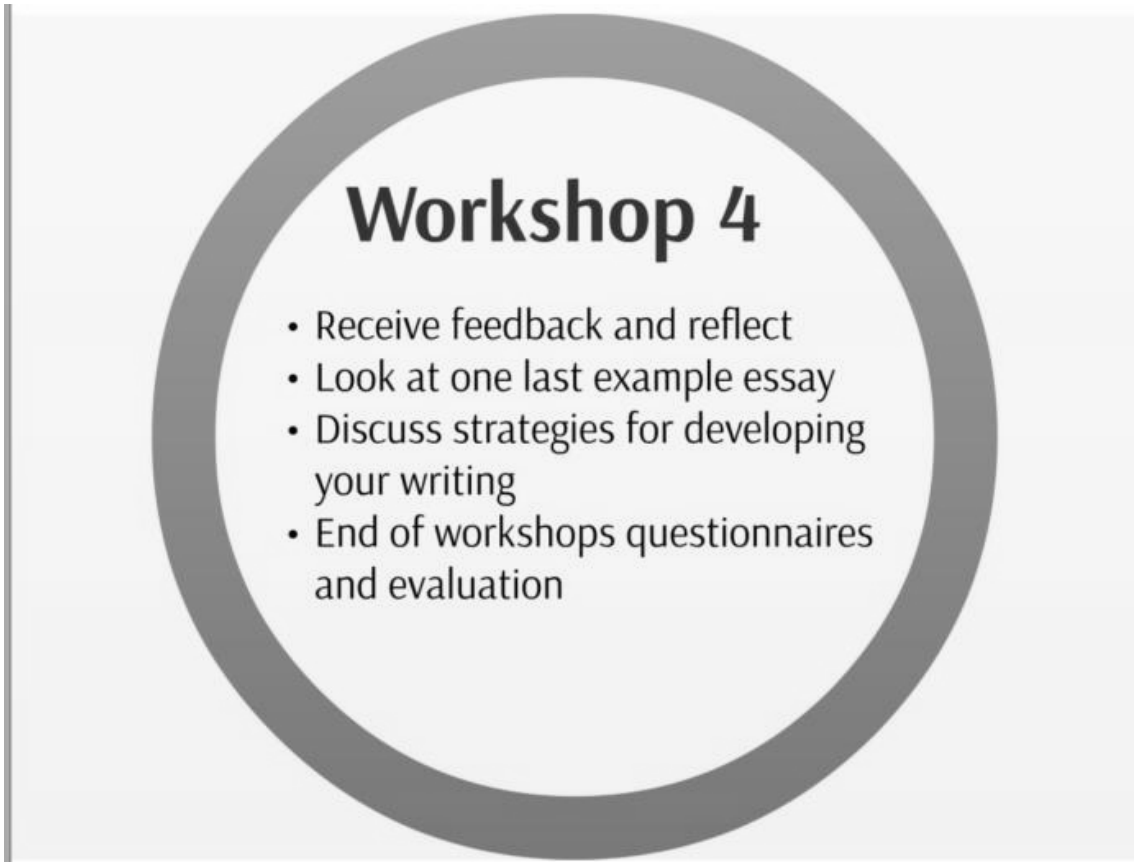
2. As you were reading the other students' essays, did you notice or recognise anything that you do in your own writing? If so, please explain.

3. **Re-read your own essay.** In the light of the reading and commenting you've done today, do you notice anything you hadn't spotted before about your own writing?

4. In the light of the reading and commenting you've done today, what's your assessment of your own essay? What do you like about it? What are its strengths? And what are its limitations, or the aspects which could be developed further?

5. Thinking about the approach you took to this essay, would you try anything different next time you write one? If so, what and how?

Workshop 4



Receiving feedback

- It's always good to take a critical approach when you read feedback. It's often assumed that all feedback is good feedback, but it's worth reading closely and considering the evidence the feedback-giver is drawing on. It's also worth weighing up the impact that taking on board such suggestions would have on your work - i.e. is it worth doing as they suggest?
- Remember that the feedback today is coming from other first year English students, so they won't (yet) have the level of expertise of tutors. But receiving feedback from three reviewers should help you put individual comments in context.

Practicalities

In your pack you should have:

- Feedback sheet
- A copy of your essay
- A reflective handout

Feel free to annotate your feedback as you read and respond to it - I'd love the yellow copy of this sheet

And the handout should help you critically reflect on the feedback you've been given (I'd like a copy of this one as well).

Workshop 4: Reflecting on feedback

1. What's your response to the comments? Do you agree with them?

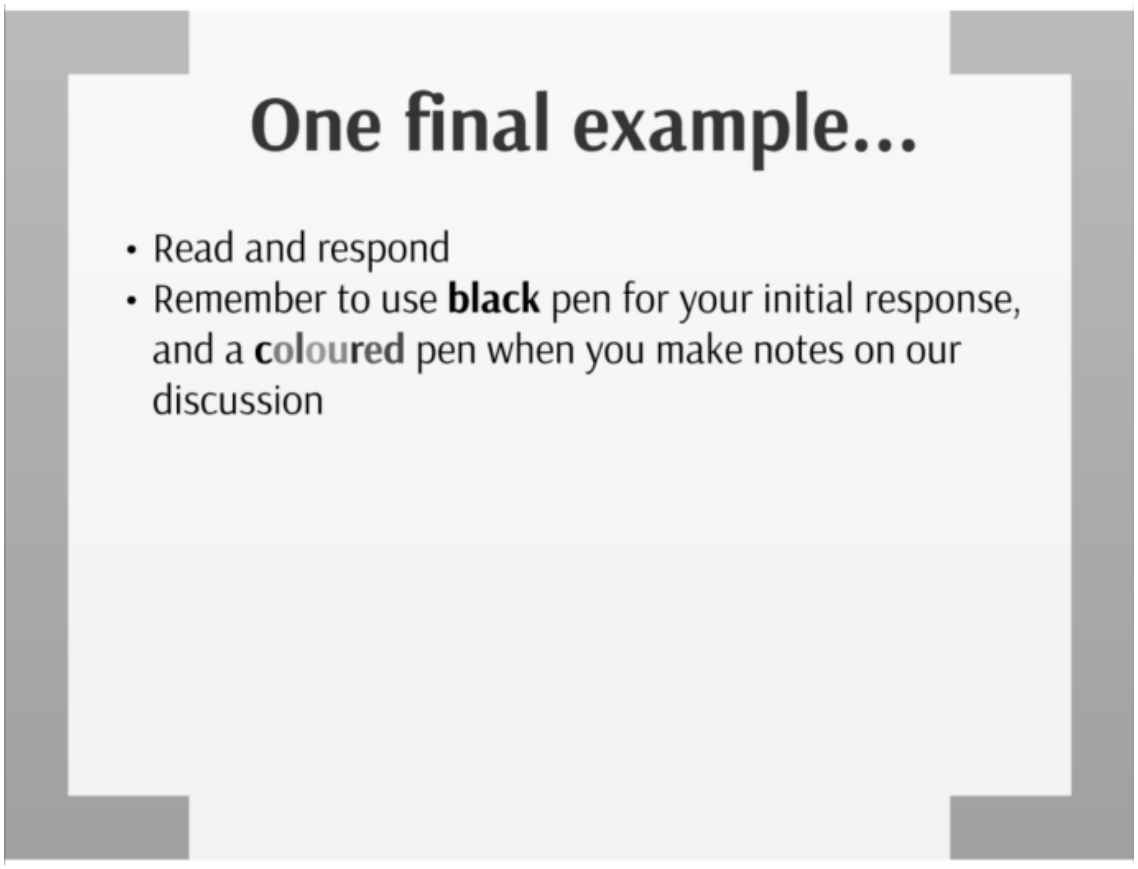
2. Which reviewer gave the most helpful comments and why?

3. Were any of the comments unhelpful? If so, why?

4. Did the reviewers see anything in your writing that you hadn't noticed before? If so, is this worth noticing?

5. Did anything surprise you about the feedback you got? Or about the three reviewers' perspectives on your writing?

6. Which of the suggestions would you consider acting upon next time you write an exam essay?



Putting all this into practice...

- What strategies do you have for developing your writing?

Next steps

Collections + one practice essay (after workshops)

- Please send in your **Collections** from this term and **at least one practice essay** written for revision **after** the workshops.
- I'll copy them and send them straight back. If you can't spare them during revision, then the end of term is fine (I can always refund postage and post essays back to your home address).

If you did a think-aloud, essay and interview in March...

- We can talk through your think-aloud transcript, and you can have a copy - this might give you some insights into your pre-writing processes (e.g. how much time you spent on planning, etc).
- It would be really really useful to the research if we could repeat the think-aloud, essay and interview before the end of term (e.g. in Weeks 6/7). It would take about 1h45 this time. Again, I can pay you for your time (£20).

Final questionnaires

- A questionnaire about writing beliefs
- A questionnaire about your prior experience with formative assessment activities like peer assessment
- An evaluation of the workshops
- Some background info if we didn't meet in March

Thank you!
Please drop me an email if you have any questions

Evaluating the workshops

I'd like to hear your views on the usefulness of these workshops. Student perspectives will be taken into account when planning any future workshop series. As with all the other data, your name will not be associated with your responses.

1. What do you feel you have learned from these workshops?

2. What do you think worked well about these workshops?

3. Which activity did you find most useful? Please circle or underline *one*.

Reviewing examples and peer essays

Receiving feedback from peers

I found both equally useful

Please explain:

4. What's your view of the holistic approach to assessing essays?

PLEASE TURN OVER

5. What was challenging or difficult about these workshops?

6. What was surprising?

7. Was there anything that you expected to learn about in these workshops, but we didn't cover?

8. What didn't work so well or could be changed in the future?

9. Any other comments?

If you think of any other comments later, please feel free to email me at:

natalie.usher@education.ox.ac.uk

Thank you!

Questionnaire on previous experience with formative assessment activities

I'd like to invite you to complete a short questionnaire about your previous experience with formative assessment activities such as peer assessment. This questionnaire will help me to report the research and to evaluate the workshops. As with all the other data you are contributing, the questionnaire will be confidential. The responses will be typed up and anonymised, and referred to only by pseudonym. The data will be dealt with and stored in exactly the same way as the other data you have contributed to the study. If you have any questions, please ask.

If you choose to fill out and return this questionnaire, this will be interpreted as an indication of your agreement to participate. In no way does this waive your rights as a participant, and you are free to withdraw from the study at any time.

Name: .....

Date: .....

Note: This page will be detached from the questionnaire and a pseudonym assigned to your responses.

Questionnaire on previous experience with formative assessment activities

Since you've been studying **at Oxford**, have you done any of the following activities **in classes or tutorials**? If yes, please briefly give some details. For 'essay', you can also think about commentary, or other forms of academic work written for tutors.

1. Taken part in a peer-assessment activity in a class or tutorial	Yes	No	Details:
2. Read example essays given to you by a tutor	Yes	No	
3. Discussed example essays with a tutor or other students	Yes	No	
4. Taken part in a self-assessment activity in a class or tutorial	Yes	No	
5. Discussed example essays with a tutor in a class or tutorial	Yes	No	
6. Self-assessed your own work by looking at the examination criteria in the handbook	Yes	No	
7. Self-assessed your own work by writing a reflection or comments to yourself	Yes	No	
8. Other similar activity (please give details)	Yes	No	

Thinking about your study at Oxford **outside of teaching time**, have you done any of the following activities? If yes, please briefly give some details. For 'essay', you can also think about commentary, or other forms of academic work written for tutors.

9. Read the English Faculty's marking or assessment criteria	Yes	No	Details:
10. Read a friend's essay or written work	Yes	No	
11. Given verbal or written comments on a friend's essay	Yes	No	
12. Shared an essay with a friend	Yes	No	
13. Received verbal or written comments on an essay from a	Yes	No	

friend			
14. Read Examiners' Report(s) for Prelims Papers 3 and 4	Yes	No	

Thinking about **the last two years of your education <u>before</u> you came to Oxford**, did you do any of the following activities in your **English** classes? If yes, please briefly give some details. You can circle both of the 'yes' boxes if applicable. If you didn't study English right before you came to Oxford, please indicate this and think about the courses where you did the most essay writing.

15. Used the exam board (or official course) criteria to self-assess your work	Yes – grade or mark	Yes – comments	No	Details:
16. Used the exam board (or official course) criteria to peer-assess another student's work	Yes – grade or mark	Yes – comments	No	
17. As a class, put together a list of criteria for a successful written task	Yes		No	
18. Used a list of criteria put together by the class to self-assess your own work	Yes – grade or mark	Yes – comments	No	
19. Used a list of criteria put together by the class to peer-assess another student's work	Yes – grade or mark	Yes – comments	No	
20. Self-assessed your work without looking at any list of criteria	Yes – grade or mark	Yes – comments	No	
21. Peer-assessed another student's work without looking at any list of criteria	Yes – grade or mark	Yes – comments	No	

CONFIDENTIAL
Demographic information

This data will be kept separately from the writing workshop data. I will only use this data to summarise who took part in the workshops. I will not do so in a way that would make it possible to identify you.

1. Name:

2. Age in May 2016:

3. Gender:

4. Ethnicity:

5. Which languages do you speak? Please specify your level of fluency and which language(s) you speak at home.

6. Where did you do most of your schooling?

Please circle: UK / other country: .....

7. What type of school did you go to?

Please circle: State comprehensive / state selective / independent selective / independent non-selective / international

Please circle: Mixed or co-ed / single sex

8. Which degree course are you taking?

.....

9. If you are doing Joint Honours, which English papers are you taking for Prelims?

.....

Appendix C:

Self efficacy questionnaire and further results

(The questionnaire was completed at the beginning of Workshop 1 and the end of Workshop 4 in both 2015 and 2016)

Writing beliefs questionnaire

I'd like to invite you to complete a short questionnaire about your beliefs about academic writing. This will help me to report the research and to evaluate the workshops. As with all the other data you are contributing, the questionnaire will be confidential. The responses will be typed up and anonymised, and referred to only by pseudonym. The data will be dealt with and stored in exactly the same way as the other data you have contributed to the study. If you have any questions, please ask.

If you choose to fill out and return this questionnaire, this will be interpreted as an indication of your agreement to participate. In no way does this waive your rights as a participant, and you are free to withdraw from the study at any time.

Name: .....

Date: .....

Note: This page will be detached from the questionnaire and a pseudonym assigned to your responses.

Writing beliefs questionnaire

Thinking of your reading, revision and writing for **Prelims Papers 3 and 4**, please rate how confident you are that you can complete the tasks described below. Please indicate your level of confidence by recording a number ranging from 0 (i.e. complete lack of confidence – you don't feel you can do this at all) to 100 (i.e. total confidence).

0	10	20	30	40	50	60	70	80	90	100
Cannot do at all					Moderately certain can do					Highly certain can do

	Choose any number from 0-100
1. If the question fits my subject knowledge, I can write a good Prelims essay	
2. I can write and finish a Prelims exam essay within the hour allowed	
3. I can use the quotation stated in the question to write my exam essay	
4. I can plan my exam essay in a way that helps me write the rest of it	
5. I can effectively manage my time over the three hours of the exam paper	
6. I understand what makes a good Prelims essay	
7. I can manage my emotions during exam writing	
8. In exams, I can plan in a way that helps me generate ideas before I start writing	
9. I understand what is expected of me in a Prelims exam essay	
10. I can write an exam essay that addresses the question or prompt	
11. I can stay focussed for the three hours of the exam	
12. I can plan my exam essay in a way that helps me organise what I'm going to say	
13. I understand what makes a good tutorial essay	
14. I can write a tutorial essay that meets my tutor's expectations	
15. After a period of revision, I can find at least three questions on the exam paper that fit my subject knowledge	
16. I know what to do with the quotation part of exam questions	
17. I can organise my essay so that it is easy for a reader to follow	
18. I can write a Prelims exam essay that meets my tutor's expectations	

Appendix C: Self-efficacy questionnaire results

All 21 students completed a self-efficacy questionnaire at the beginning of Workshop 1 and the end of Workshop 4. One student was excluded from the analysis as while he completed the course, he only attended one group session due to illness.

The data were analysed to address RQ4: What is the impact of participating in a peer assessment intervention on students' self-reports of self-efficacy for writing? To guide statistical analysis, this RQ is rephrased as: is there a significant difference in students' ratings of self-efficacy for each item at Workshop 1 and Workshop 4? The null hypothesis is that there is no difference in students' ratings of self-efficacy at Workshop 1 and at Workshop 4 on each item.

Initial descriptive analysis of group means and mean differences, as shown in Table 1, shows that for all items, there was quite a lot of variance in how students used the 100-point scale to rate their self-efficacy. The standard deviations are quite high. For many items, the mean differences are over 10 points, with the group mean difference for item 6, 'I understand what makes a good Prelims essay', particularly large at 23.5 points. The smallest mean differences are for items which could be considered 'control' items: 'I understand what makes a good tutorial essay' at 0.25 points, and 'I can write a tutorial essay that meets my tutor's expectations' at 2.55 points.

Table 1: Group means and mean differences

		Group mean W1	Std. deviation	Group mean W4	Std. deviation	Group mean diff.	Std. deviation
1	If the question fits my subject knowledge, I can write a good Prelims essay	71.00	16.19	80.00	10.39	9.00	14.47
2	I can write and finish a Prelims exam essay within the hour allowed	65.50	17.61	76.75	14.17	11.25	12.66
3	I can use the quotation stated in the question to write my exam essay	62.50	19.30	79.65	11.35	17.15	15.21
4	I can plan my exam essay in a way that helps me write the rest of it	58.80	18.38	72.25	14.37	13.45	17.76
5	I can effectively manage my time over the three hours of the exam paper	63.00	21.48	71.00	18.32	8.00	10.69
6	I understand what makes a good Prelims essay	61.50	15.05	85.00	6.69	23.50	13.09

7	I can manage my emotions during exam writing	75.50	17.01	79.65	14.62	4.15	13.44
8	In exams, I can plan in a way that helps me generate ideas before I start writing	64.70	19.23	72.05	12.81	7.35	14.43
9	I understand what is expected of me in a Prelims exam essay	66.00	12.10	83.00	8.18	17.00	12.07
10	I can write an exam essay that addresses the question or prompt	70.45	14.34	79.75	11.18	9.30	13.49
11	I can stay focussed for the three hours of the exam	69.65	21.40	73.10	17.86	3.45	12.45
12	I can plan my exam essay in a way that helps me organise what I'm going to say	66.25	19.93	70.00	13.48	3.75	18.84
13	I understand what makes a good tutorial essay	75.80	14.40	76.05	12.61	.25	10.64
14	I can write a tutorial essay that meets my tutor's expectations	72.40	12.75	74.95	11.50	2.55	10.65
15	After a period of revision, I can find at least three questions on the exam paper that fit my subject knowledge	77.50	16.98	82.00	12.29	4.50	10.12
16	I know what to do with the quotation part of exam questions	67.00	20.16	81.100	15.03	14.10	17.80
17	I can organise my essay so that it is easy for a reader to follow	64.90	19.10	74.750	11.41	9.85	15.29
18	I can write a Prelims exam essay that meets my tutor's expectations	60.25	17.88	71.875	9.14	11.63	14.20

To investigate whether these mean differences are statistically significant, I explored the data to see whether it would meet the assumptions necessary for a t-test. As according to Field (2013), it is important that the mean differences are distributed normally, I explored these visually using histograms and P-P plots and then with tests of normality. As shown in Table 2, most of the items deviated from the normal distribution on their mean differences. This is unsurprising, given the small sample size. Therefore, I decided to use a non-parametric test for all items – it would not be meaningful to use the t-test for some items but not others.

Table 2: Exploring normality: mean differences

		Group mean difference	Std. deviation	Skewness	Std. error	Kurtosis	Std. error	Kolmogorov-Smirnov		Shapiro-Wilk	
								Statistic	Sig.	Statistic	Sig.
1	If the question fits my subject knowledge, I can write a good Prelims essay	9.00	14.47	**1.205**	.512	.677	.992	.222	**.011**	.849	**.005**
2	I can write and finish a Prelims exam essay within the hour allowed	11.25	12.66	.704	.512	-.010	.992	.189	.058	.931	.160
3	I can use the quotation stated in the question to write my exam essay	17.15	15.21	**1.293**	.512	2.489	.992	.176	.107	.898	**.037**
4	I can plan my exam essay in a way that helps me write the rest of it	13.45	17.76	.518	.512	-.797	.992	.176	.107	.925	.124
5	I can effectively manage my time over the three hours of the exam paper	8.00	10.69	.763	.512	.386	.992	.226	**.009**	.908	.058
6	I understand what makes a good Prelims essay	23.50	13.09	.677	.512	-.218	.992	.142	.200*	.927	.137
7	I can manage my emotions during exam writing	4.15	13.44	.400	.512	-.814	.992	.193	**.050**	.923	.112
8	In exams, I can plan in a way that helps me generate ideas before I start writing	7.35	14.43	**1.416**	.512	2.680	.992	.220	**.012**	.869	**.011**
9	I understand what is expected of me in a Prelims exam essay	17.00	12.07	.099	.512	-.011	.992	.202	**.032**	.929	.150
10	I can write an exam essay that addresses the question or prompt	9.30	13.49	-.120	.512	1.312	.992	.236	**.005**	.922	.111

		Group mean difference	Std. deviation	Skewness	Std. error	Kurtosis	Std. error	Kolmogorov-Smirnov		Shapiro-Wilk	
								Statistic	Sig.	Statistic	Sig.
11	I can stay focussed for the three hours of the exam	3.45	12.45	-.631	.512	-.150	.992	.200	**.036**	.931	.165
12	I can plan my exam essay in a way that helps me organise what I'm going to say	3.75	18.84	-.066	.512	-.743	.992	.130	.200*	.970	.760
13	I understand what makes a good tutorial essay	.25	10.64	-.499	.512	-.135	.992	.191	.055	.933	.173
14	I can write a tutorial essay that meets my tutor's expectations	2.55	10.65	-.129	.512	-1.073	.992	.158	.200*	.952	.406
15	After a period of revision, I can find at least three questions on the exam paper that fit my subject knowledge	4.50	10.12	.828	.512	.701	.992	.222	**.011**	.927	.138
16	I know what to do with the quotation part of exam questions	14.10	17.80	.499	.512	.502	.992	.170	.132	.944	.289
17	I can organise my essay so that it is easy for a reader to follow	9.85	15.29	.367	.512	.176	.992	.146	.200*	.965	.638
18	I can write a Prelims exam essay that meets my tutor's expectations	11.63	14.20	**1.152**	.512	1.705	.992	.196	**.044**	.918	.092

*According to SPSS, this is a lower bound of the true significance.

Thus, I ran a Wilcoxon signed-rank test to compare students' ratings of self-efficacy on each item. As shown in Table 3 below, the null hypothesis that is that there is no difference in students' ratings of self-efficacy at Workshop 1 and at Workshop 4 can be rejected for the following items if we take the p-value at 0.05: Item 1, Item 2, Item 3, Item 4, Item 5, Item 6, Item 8, Item 9, Item 10, Item 16, Item 17, Item 18. These items are discussed below, together with their effect sizes.

I then calculated the effect size for each item according to the equation in Field (2013, p. 227):

$$r = \frac{z}{\sqrt{N}}$$

In this equation, *z* is the z-score generated by SPSS (and included in the table below). *N* is the number of observations on which *z* is based, so 20. Field recommends using Cohen's guidelines for interpreting the effect sizes, so as you can see in the table below, several items have large effect sizes of over .70. If we were to apply Hattie's benchmarks for effect sizes in education, these would be very large indeed. However, I think these effect sizes should be interpreted with particular caution as the study is based on small sample and so does not have much statistical power. Even for non-significant results with small mean differences there are small effect sizes. For example, Item 15, 'after a period of revision, I can find at least three questions on the exam paper that fit my subject knowledge' has an effect size of $r = 0.42$ despite there being no significant difference between self-reports at Workshop 1 ($Mdn = 80.00$) and Workshop 4 ($Mdn = 80.00$; $z = 1.870$; $p = .062$).

Reflecting on today's writing process

These open-ended questions are to help you reflect on your writing process today. Your responses **won't** be shared with the students who will read and assess your essay

1. Describe the question you chose. What was it asking you to do?

2. Why did you choose this question?

3. Did you try anything new today? If so, what? And how did it work out?

4. Briefly describe what you did if/when you planned.

5. As you wrote, did your original intentions or plans for the essay change? For example, did the way that you saw the question change? Did new ideas come to mind?

6. What went well about this essay?

7. What did you find challenging about this essay?

8. Thinking about the approach you took to today's essay, would you change anything about it next time?

9. Did you do any revision or preparation for today's essay? If so, please give brief details (e.g. topic/authors, revision techniques)

10. Anything else what comes to mind about the process of writing today's essay?

Writing Processes Questionnaire

Please think about the essay you've just written as you respond to this questionnaire.

Rate the extent to which <u>you intended to do</u> each of the following from 1 (not at all) to 4 (yes).

		1	**2**	**3**	**4**
1.	Thoroughly understand the question.	Not at all	Not really	Sort of	Yes
2.	Make good quality goals or plans.	Not at all	Not really	Sort of	Yes
3.	Monitor my progress.	Not at all	Not really	Sort of	Yes
4.	Evaluate my progress.	Not at all	Not really	Sort of	Yes
5.	Adapt my learning or approach if necessary.	Not at all	Not really	Sort of	Yes

Thinking about writing the essay you described above, rate the extent to which <u>you actually did</u> each of the following from 1 (not at all) to 4 (yes).

Before *I started writing the first full line of my essay, I….*

		1	**2**	**3**	**4**
6.	Reflected about what the question was asking me.	Not at all	Not really	Sort of	Yes
7.	Considered what knowledge I was being asked to show.	Not at all	Not really	Sort of	Yes
8.	Thought about why I am being asked to know this stuff.	Not at all	Not really	Sort of	Yes
9.	Identified what skills I need to demonstrate.	Not at all	Not really	Sort of	Yes
10.	Made sure I understood terminology used in the question.	Not at all	Not really	Sort of	Yes
11.	Thought about my tutor's expectations for the essay.	Not at all	Not really	Sort of	Yes
12.	Considered what knowledge or ideas I should demonstrate.	Not at all	Not really	Sort of	Yes
13.	Reflected on why this task is important.	Not at all	Not really	Sort of	Yes
14.	Thought about what resources I should use for this task (notes, readings, texts).	Not at all	Not really	Sort of	Yes
15.	Set goals for my work.	Not at all	Not really	Sort of	Yes
16.	Decided how I would use my time.	Not at all	Not really	Sort of	Yes
17.	Identified which content and ideas I could use.	Not at all	Not really	Sort of	Yes

18. Considered feedback I had received on previous work.	Not at all	Not really	Sort of	Yes
19. Generated some ideas about what I could write about.	Not at all	Not really	Sort of	Yes
20. Organised or structured my knowledge to help me write the essay.	Not at all	Not really	Sort of	Yes
21. Made a plan that could be completed within the 1 hour limit.	Not at all	Not really	Sort of	Yes
22. Organised what I was going to say to make it easier for a reader to follow.	Not at all	Not really	Sort of	Yes

During or shortly after *writing the essay, I…..*

	1	**2**	**3**	**4**
23. Asked myself if I was understanding the task.	Not at all	Not really	Sort of	Yes
24. Reflected on the way I was doing or did the task.	Not at all	Not really	Sort of	Yes
25. Checked to see if I was sticking to time.	Not at all	Not really	Sort of	Yes
26. Considered whether my plan was appropriate for the essay question.	Not at all	Not really	Sort of	Yes
27. Considered my feelings for the task.	Not at all	Not really	Sort of	Yes
28. Evaluated the level of effort I was putting in.	Not at all	Not really	Sort of	Yes
29. Asked myself if I was answering the question.	Not at all	Not really	Sort of	Yes
30. Thought about feedback I had received on previous work.	Not at all	Not really	Sort of	Yes
31. Made a judgement about the usefulness or value of the ideas I was using.	Not at all	Not really	Sort of	Yes
32. Evaluated how well I was answering the question.	Not at all	Not really	Sort of	Yes
33. Changed my understanding of what the question was about.	Not at all	Not really	Sort of	Yes
34. Modified my plans for the essay.	Not at all	Not really	Sort of	Yes
35. Switched to a different strategy or approach.	Not at all	Not really	Sort of	Yes
36. Changed my feelings about the task.	Not at all	Not really	Sort of	Yes
37. Altered the level of effort I was putting in.	Not at all	Not really	Sort of	Yes
38. Modified my beliefs about how well I would do on this task.	Not at all	Not really	Sort of	Yes

Think-aloud, writing and interview schedule

1. Explain what we'll do today

First, I'll ask you to think aloud. You'll have some practice on some word conundrums, and then you'll keep thinking aloud as you go through the paper, and choose your question. Next, you'll write the essay without thinking aloud. Finally, once you've finished writing, I'll ask you to look back on the essay while it's still fresh. I'll also ask you some interview questions about writing more generally.

The **reason** I'm asking you to do this is because a focus of my research is essay writing processes. As the writing process unfolds, writers often forget what they were thinking about or what they intended to do at the beginning.

It's really **not a test**.

Which exam paper would you like to do?

While you're writing, I'll stay in the room, but I'll just be reading an article

Copy of the essay: I will give you a photocopy of the essay you wrote so you can use it for revision if you want. I'll also give you a photocopy of the essay you brought along today. Sometimes people find reading their think-aloud transcripts interesting or insightful. If you would like, once the workshops are over, I can give that to you and we can discuss it.

2. CONSENT AND PAYMENT

Give the student the opportunity to read the letter and consent form. Make tea!

Stress that s/he can withdraw at any time without any consequences - free to stop at any time, payment does not affect this.

Participant preference on pseudonym?

3. THINK-ALOUD PRACTICE

In a moment, I will give you a poem and some questions about it to practice thinking aloud. As you read, think to yourself, or write, please read and think aloud, even as you are writing something down. Please don't explain or justify what you are doing. If you

fall silent for more than 10 seconds, I will remind you to talk by saying 'keep talking.'

Otherwise, please ignore me as much as possible.

PRE-WRITING THINK-ALOUD

> In a moment, I will give you an exam paper. Please read through it and pick one essay to write today. Work on this task as you normally would – read, think, write down notes, highlight. As you do this, please think aloud – as if you were talking to yourself. You don't need to talk about your processes or to explain or justify what you are doing. Please just focus all your attention on doing the task. Please continue to think aloud even when you are reading and writing. When you get to the point where you are about to start writing the essay in full sentences, you don't need to think-aloud. I'm going to give you 1hr 10 mins in total to do your thinking-aloud and write the essay.

4. WRITING

Give time check 5 minutes before the participant is due to finish.

5. POST WRITING INTERVIEW

The essay

- How did it go?
- Tell me about how you came to choose this question
    - How confident were you about the topic?
- Talk me through what you did after you stopped thinking aloud.
    - What did you do after that?

[If the student didn't do the think-aloud, ask more specific questions, using the jottings on the exam paper and the planning for stimulated recall – what were you thinking about when you were writing your plan?]

- Did your approach differ to **how you've done timed essays in the past?**
- Were you able to use any ideas that you've written about before?
- Did any new ideas pop up as you were writing?
- Did the way you saw the question and/or quotation change as you were writing?

- What do you think the strengths of this piece of writing are?
- If you had ten more minutes to work on it, what would you do with that time?
- Would you do anything differently next time?
- Were you thinking of a particular audience or reader as you were writing?
- For this particular question, what is the question asking you to do?
    - If necessary, stress: this particular question
    - Or: what are the question setters asking you to do in this question?
    - Or: what is the purpose of this particular question?
- **Tell me about the revision you've been doing. Did you do any revision for today's essay? [Prompt to explain]**

Comparing essays; quality

- **If you compare the essay you wrote during the workshops and the one you wrote today, which do you think is stronger? Can you explain why?**
- **For you, what makes a good exam essay for Prelims?**
    - Elicit: what's the role of critics in your writing?
    - Or: how do you use quotations?
    - Or: how are you expected to use quotations?
- What do you think makes a good tutorial essay?
- For you, what is the purpose of writing an exam essay?
    - Rephrase as: what are the examiners expecting you to do?
- Is the purpose of writing an exam essay different to writing a tutorial essay?

A-level / prior writing experiences

- [Post: Last time we talked about about your experience with writing at A-level / before you came to Oxford. Thinking about that…] [Pre-: Ask what students were studying before they came to Oxford]. Thinking about your writing before you came to Oxford / thinking about A-level, what strategies and approaches do you still find useful when you write an Oxford exam essay?
    - Or: what can you draw on from your experience before coming to Oxford?
    - Or: how is this type of essay writing similar to the writing you do before coming to Oxford? (e.g. A-level).

    - Or: ask about specific approaches or strategies that students talk about – e.g. how/when did you develop that approach / that strategy?

- And how do you think this type of essay is different to A-level / the type of essays you wrote before coming to Oxford?
    - Elicit here: in terms of the **purpose** of the task?
    - Or: in terms of what the task asks you to do?

(Post-only) Workshops and writing

- What did you learn in the workshops?
    - Have you had the chance to put this into practice yet?
- What did you think of the feedback you got during the workshops? [bring a copy to discuss].
    - Have you been able to use this feedback in your writing?
- Is there anything that you learned about in the workshops that you wanted to us but haven't been able to put into practice?
    - What are the barriers or things that make it difficult to do these things?
- What did you think of the strategies handout?
    - Have you been able to use the strategies handout?

Closing:

- Final question: do you think that thinking aloud at the beginning changed the way you wrote your essay?
- Anything else you'd like to add, or any questions?

(Post-only) Ask student if they'd like to go through their think-aloud from February.

University of Oxford
CENTRAL UNIVERSITY RESEARCH ETHICS COMMITTEE (CUREC)
CUREC /1A CHECKLIST FOR THE SOCIAL SCIENCES AND HUMANITIES

The University of Oxford places a high value on the knowledge, expertise, and integrity of its members and their ability to conduct research to high standards of scholarship and ethics. The research ethics clearance procedures have been established to ensure that the University is meeting its obligations as a responsible institution. They start from the presumption that all members of the University will take their responsibilities and obligations seriously and will ensure that their research on human subjects is conducted according to the established principles and good practice in their fields and in accordance, where appropriate, with legal requirements. Since the requirements of research ethics review will vary from field to field and from project to project, the University accepts that different guidelines and procedures will be appropriate. Please check the CUREC website to ensure that you have the correct checklist for your project.

ONLY TYPE-WRITTEN CHECKLISTS WILL BE ACCEPTED. PLEASE DO NOT SEND HANDWRITTEN CHECKLISTS.
ONLY EMAILED APPLICATIONS WILL BE ACCEPTED. PLEASE DO NOT SEND APPLICATIONS BY POST.

What this checklist will not assess

This checklist does not cover research governance, satisfactory methodology, or compliance with the requirements of publishers when administering their tests or questionnaires. As principal researcher, it is your responsibility to ensure that requirements in these areas are met.

CUREC does not review studies classed as *audit* (see glossary and Decision Flowchart for CUREC on our website).

The use of an *asterisked word underlined* in this checklist indicates a phrase defined in CUREC's glossary. The glossary and further information on the University's research ethics procedures are available from the CUREC website: www.admin.ox.ac.uk/curec

What this checklist is designed for

This CUREC /1A checklist is designed largely for research that falls within the Divisions of Social Sciences and Humanities where ethical issues are relatively few and straightforward. Interviews, field work and oral history are also included in the CUREC process.

The full CUREC /2 application is only required where certain project characteristics (eg type of participants, or procedures) result in a more complex set of ethical issues. It is expected that only in a limited number of cases will it be necessary for researchers to complete a CUREC /2 application. The checklist below will direct you to a CUREC /2 application if needed.

Office use only: IDREC Ref. No. _____

Date of confirmation that checklist accepted on behalf of IDREC: // //

Please complete the sections that follow and follow prompts to stop completion and/or submit other documents. ***Please indicate your answer to all the Yes / No questions with a type-written "X".***		
Example: ***1) By taking part in the research, will participants be at serious risk of criminal prosecution (e.g. by providing information on drug abuse or child abuse)?***	**YES** **X**	**NO**

SECTION A: filter for CUREC /2 application.

This section determines whether your study raises more complex issues which require the completion of a full application for ethical review, known as the CUREC /2 application.		
1) Are research participants classed as "*people whose ability to give free and informed consent is in question*"? (This may include those under 18 (though see FAQ C12), prisoners or vulnerable adults) *Note: sometimes participants aged 16-17 are regarded as "competent youths" and may be able to consent for research themselves. Please consult the FAQ C12 on the CUREC website. However, if any of your participants are aged 16 or under, please answer 'yes' here and also answer question 4) below.*	**YES**	**NO** **X**
2) By taking part in the research, will participants be at serious risk of criminal prosecution (e.g. by providing information on drug abuse or child abuse)?	**YES**	**NO** **X**
3) Does the research involve the *deception* of participants?	**YES**	**NO** **X**
If you answered NO to all of the questions above please go to SECTION B. If you have answered yes to any question above continue to question 4 below.		
4) Is your project covered by one of our approved research protocol(s)? (see CUREC website) *If yes, please give protocol number(s):*	**YES**	**NO**
If you answered Yes to ANY of questions 1-3, and answered No to question 4, please STOP COMPLETING THIS CHECKLIST and do not submit it for ethical review. ***Instead, please complete the CUREC/2 application form, available to download from the CUREC website. Then submit the CUREC/2 for ethical review.*** ***If you answered Yes to ANY of questions 1-3, and answered yes to question 4, please go on to SECTION B.***		

SECTION B: contact details and project description (NB must be typed not handwritten).

Contact details	
1) *Principal researcher* / supervisor (if student research):	Dr Therese Hopfenbeck Dr Kathleen Quinlan
2) Name of student (if student research):	Natalie Usher
3) Degree programme, e.g. DPhil, BA, MPhil, BSc, MSc (if student research):	DPhil Education
4) Department or institute name:	Education
5) Address for correspondence (if different from 4):	
6) University email address and phone contact:	therese.hopfenbeck@education.ox.ac.uk; 01865 274017 kathleen.quinlan@learning.ox.ac.uk ; 01865 286824 natalie.usher@education.ox.ac.uk ; 01865 284098
7) Name and status of others taking part in the project. E.g. 3rd year undergraduate; post-doctoral research assistant:	
Project description	
8) Title of research project:	Learning about academic writing through holistic peer review
9) List location(s) where project will be conducted:	Oxford
10) If your research involves overseas travel or fieldwork, by the time the research starts, will you have completed and returned a travel risk assessment form? (This may be necessary to ensure that the activity is covered by the University's travel insurance – see http://www.admin.ox.ac.uk/finance/insurance/travel/)	N/A
11) Anticipated duration of project:	21 months
12) Anticipated start date:	5/1/2015
13) Anticipated end date:	30/9/2016

14) Title and brief lay description of *research* (about 150 words), plus description (about 200 words) of the nature of participants.

When describing the research, please include your methodology, how you are applying professional guidelines, and the use to which results/data will be put. Please also declare any conflicts of interest here.

When describing participants, please include criteria for inclusion/exclusion, method of recruitment, and processes (as appropriate) for consent to participate. To support your description please attach documentation such as:

- *recruitment and advertisement material eg a poster or invitation letter*
- *information for participants to read before they agree to take part eg participant information sheet(s)*
- *a document to record informed consent eg written consent form(s) or oral consent script in case of an oral consent process*
- *a guide to interview questions (this may be a list of questions to be asked, or a preliminary scope of questions) or sample of other instruments such as a sample questionnaire*
- *debriefing document after participants have taken part (if relevant)*

Detailed guidance is found on the CUREC website

Title and brief lay description of research (150 words)
The study is titled 'Learning about academic writing through peer review'. The data collected will be analysed to answer the research questions, which focus on the impact of participating in a peer assessment intervention on a) students' knowledge about quality in disciplinary writing, b) their judgments about their own writing and others', c) the ways they interpret tasks, and d) their writing achievement.

The peer assessment intervention will be evaluated through the following methods: 1. Pre- and post-intervention think-aloud activities (students think aloud about a task, write a timed essay and participate in a retrospective interview). 2. A pen-and-paper evaluative expertise assessment. 3. Written responses completed for intervention activities, e.g. peer feedback. 4. Responses to diary prompts. 5. Essays written before, during and after the intervention. 6. open-ended questionnaire items

Before the intervention, a small corpus of essays will be assembled. Students will be asked to submit work to this for use in the intervention and for university teaching purposes. Some of these samples will be discussed by tutors in interviews about text quality.

Description of participants and obtaining informed consent (200 words)
There are four possible groups of participants: 1. students contributing written work to the corpus (20-30 participants), 2. students participating in the pilot activities only (approx. 3 participants), 3. students participating in the full intervention, including pre- and post- intervention activities (12-24 students), 4. college English tutors participating in interviews (approx. 3 participants)

Those eligible to participate in activities 1, 2 and 3 are students taking the undergraduate English or joint honours course at the University of Oxford. The intervention participants will be in their first year of study. The recruitment strategy will negotiated with the faculty. Either an open call will be sent through the faculty, or more likely, I will be introduced to the college tutors, who will be asked to forward students an introductory email.

Group 1 will be sent an information sheet before they submit the work, and given the opportunity to ask questions. Participants will be asked to submit written consent with the work. The consent process may be face-to-face, via email or via university post.

Groups 2, 3 and 4 will be given an information sheet to read and the opportunity to ask questions at the beginning of the research encounter. Participants will then be asked to give written consent.

At a later date, tutors will be recruited to rate the quality of the essays in the study. I will complete a separate application nearer the time.

15) What are the ethical issues connected with your research and what steps have you taken to address them?

Please do not answer "none".

The committee wants to see evidence that you have identified potential ethical issues with respect to your research and have taken steps to address them. These issues could relate to your own safety as a researcher, participant burdens/risks, and data protection/confidentiality. For more guidance on ethical issues, please see www.admin.ox.ac.uk/curec/resources/

Recruitment of participants: I am aware that recruiting students through their college tutors or the faculty may put pressure on them to participate. I will emphasise to students that this is an extra activity that is in no way part of their course, that they are free to choose to participate, and that there is no academic or other penalty for choosing not to participate or for withdrawing from the study.

Workload burden, stress and cash payments: Undergraduate students have heavy workloads, and this study will take place at a stressful time of year, in Hilary and Trinity terms in the run-up to Prelims exams. The intervention workshops themselves are on examination writing and thus should be beneficial to students as they prepare. Nonetheless, I will be sensitive to the signs of exam stress and related issues. I will foster a supportive atmosphere in the intervention, and design interview items and other activities to avoid further stress. Students will be trained to give feedback in a constructive and considerate manner.

In view of students' heavy workloads, small payments will be made in recognition of the time given to activities outside of the intervention teaching time. Student participants will be paid £5 for submitting an essay to the corpus, £20 for participation in the pilot (approx. 2 hours) and £15 for participation in the pre-intervention think-aloud activity (1hr30mins-1hr45mins). Some participants will also take part in a post-intervention think-aloud activity, and they will be paid separately for participation for this. The amounts are in line with what has been offered in past studies in Oxford and are intended to be modest so that students are not pressured to participate.

The payments will be made in cash so the participants are free to choose how to spend them. The payment will be made at the beginning of the research encounter in order to make it clear that it is not dependent on their answers. I will emphasise that the payment does not affect students' rights as a participant and that they do not have to return it should they withdraw.

Confidentiality: The essays collected in the study will be read by other participants. The essays collected for the corpus will be read by the tutor interviewees, the students participating in the intervention, and could also be used in the future for teaching purposes. The essays written by students participating in the intervention will be read by tutors rating this work for achievement, and by other students participating in the intervention when offering feedback. Students will be made aware of these audiences for their work.

Thus all handwritten texts collected for the intervention will be typed up to prevent identification from handwriting, and all identifying information will be removed. A codes system will be used to keep track of which texts belong to which participants, and participants will have the option of choosing their own pseudonym. The document containing the key will be kept securely, as detailed below. When texts are to be shared with other participants, no identifying codes will be used (although texts may be given arbitrary labels, e.g. Text 1, to make it easier for classroom discussion).

I will make it clear to participants that I cannot be certain that these steps will guarantee anonymity.

Other users of the data: The data will be read by myself, my supervisors, transcriber(s) and one or two other research students offering research assistance as a reciprocal agreement for inter-coder agreement. The transcriber and fellow research students will be asked to sign a confidentiality agreement. I do not plan to share the data more widely nor deposit it in a data archive as this is a small-scale study.

Students receiving payment for activities will be asked to sign a receipt so that this can be reimbursed from my ESRC RTSG budget. The information on receipt will be used for the sole purpose of processing the reimbursement of the payment, and will be neither disclosed nor sent on anywhere other than the finance division. Participants will be informed of this.

16) Will you obtain informed consent according to CUREC guidelines and good practice in your discipline *before* participation?	**YES**	**NO**
	X	
If you have marked No, please give a brief explanation and justification for this decision here:		

17) Will your research involve discussing sensitive issues? *This could be information relating to race or ethnic origin, political opinions, religious beliefs, physical/mental health, trade union membership, sexual life or criminal activities.*	**YES**	**NO**
If you have marked Yes, please make sure that you have included some supporting information (as directed in qu 14 of this section) showing the range of questions covering these issues.		**X**

18) Will you ensure that *personal data* collected directly from participants or via a *third party* is held and processed in accordance with the provisions of the Data Protection Act?	**YES**	**NO**
	X	

19) How will you ensure that any personal or sensitive data are captured, transferred and stored securely?

In particular if data are to be captured electronically, please consult with your local IT department and, with respect to University IT security policies, please comment on how you will capture such data in the first instance, how you will transfer them over networks or via portable media and how and where data will be stored. For more information please see the University's web pages on research data management:
http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/

Data will be collected in handwritten or audio formats; additionally, students may e-mail work. All handwritten data will be typed up. Hard copies of data, e.g. handwritten essays, will have their identifying information removed and replace with a code. The data will be kept in a locked filing cabinet in an office at the Education department.

Audio-recorded data will captured on a voice recorder. Once transferred onto the main research computer, the audio data will be removed from the voice recorder device. When transcription services are used, the audio files will be transferred by the OxFile service rather than by email. The transcriber will be instructed not to include identifying data i.e. names in the transcription, and to delete the file on completion of the transcription.

Where data is emailed by students, this will be made into a file. The original email will be deleted.

No identifying information will be left in the digital data files. A code system will be used. All files will be named and identified using this code. One hard copy will be kept of the key, and one on the research computer in a separate folder from the data.

All digital data will be stored under password protection on my personal laptop. Backups will be made via the OUCS HFS service and on one external hard drive, which will be kept in a locked place. Backups will be made at least twice weekly.

SECTION C: methods and procedures to be used.

In this section, please mark X against as many methods or procedures as apply to your research. (This section will help the committee understand the nature of your research and may be used for audit).	
	Please mark "X"
1) Analysis of existing records	
2) Snowball sampling (recruiting through contacts of existing participants)	X (may be used for collection of texts for corpus)
3) Use of casual or local workers eg interpreters	
4) Participant observation	
5) Covert observation	
6) Observation of specific organisational practices	
7) Participant completes questionnaire in hard copy	X
8) Participant completes online questionnaire or other online task	
9) Using social media	
10) Participant performs paper and pencil task	X
11) Participant performs verbal or aural task (eg for linguistic study)	X
12) Focus group	
13) Interview	X
14) Audio recording of participant	X
15) Video recording of participant	
16) Photography of participant	
Other (please specify):	

SECTION D: professional guidelines and training.

In this section, please mark X against at least one of the following professional guidelines. You should use the principles listed in your chosen guideline(s) in conducting your own research. Note this is not an exhaustive list.		
Research specialism/methodology	**Association and guidance document (active links below)**	**Please mark "X"**
Anthropology	Association of Social Anthropologists of the UK and Commonwealth	
Criminology	British Society of Criminology: Code of Ethics for Researchers in the Field of Criminology	
Education	British Educational Research Association Ethical Guidelines for Educational Research	**X**
Geography	Association of American Geographers Statement on Professional Ethics	
Geography	Royal Geographical Society: Research Ethics and Code of Practice	
History	Oral History Society of the UK Ethical Guidelines	
Law (Socio-Legal)	Socio-Legal Studies Association: Statement of Principles of Ethical Research	
Management	Academy of Management's Professional Code of Ethics	
Political Science	American Political Science Association (APSA) Guide to Professional Ethics in Political Science (Section H)	
Politics	Political Studies Association, Guidelines for Good Professional Conduct	
Psychology	British Psychological Society Code of Ethics and Conduct • British Psychological Society: Conducting Research on the Internet: (Guidelines for internet-mediated research)	

Social Research	Social Research Association: Ethical Guidelines • Social Research Association : Data Protection Act 1998 : Guidelines for Social Research • Social Research Association: Code of Practice for the Safety of Social Researchers	
Sociology	The British Sociological Association: Statement of Ethical Practice	
Visual Research	ESRC National Centre for Research Methods Review Paper: Visual Ethics: Ethical Issues in Visual Research	
Other professional guidelines ***Please specify the other guideline(s) used here:***	ESRC Framework for Research Ethics	
Please indicate what training in research ethics the researchers involved with this study have received, e.g. the title of the course and date completed (online training available at www.admin.ox.ac.uk/curec/training/) *If no formal training has been received, please indicate any discussions of research methodology between researchers and supervisors here.*		
Foundations of Educational Research lecture and seminar (David Mills, Pam Sammons, November 2012) Conceptualising Educatonal Research seminar (Alis Oancea, November 2012)		

SECTION E: signatures.

Please ensure this checklist is signed by:

For staff research:	*For student research:*
1) *Principal researcher*	**1) *Principal researcher* (project supervisor)**
2) Head of Department (or nominee)	**2) Head of Department (or nominee)**
-	**3) Student researcher**

2) Departmental endorsement signature

I have read the research project application named above. On the basis of the information available to me, I:

(i) consider the principal researcher to be aware of her/his ethical responsibilities in regard to this research;

(ii) consider that any ethical issues raised have been satisfactorily resolved or are covered by relevant professional guidelines and/or CUREC approved protocols, and that it is appropriate for the research to proceed (noting the principal researcher's obligation to report should the design of the research change in a way which would alter any of the above responses so as to require completion of a CUREC/2 full application);

(iii) consider that the principal researcher/supervisor/student researcher is aware of her/his ethical responsibilities in regard to this research. I am satisfied that: the proposed project design and scientific methodology is sound; the project has been/will be subject to appropriate *peer review*; and is likely to contribute to existing knowledge and/or to the education and training of the researcher(s) and that it is in the *public interest*.

Signed by Head of Department or nominee:……..

(*example nominees for student research include the Director of Graduate Studies / Director of Undergraduate Studies*)

Print name (block capitals)……

Date:……

3) Student signature (if student research)

I understand the questions and answers that have been entered above describing the research, and I will ensure that my practice in this research complies with these answers, subject to any modifications made by the principal researcher properly authorised by the CUREC system.

Signed by student:…See email confirmation……..

Print name (block capitals)……NATALIE USHER…… Date:…28/11/2014…

SECTION F: final check before submitting your application.

Please use this section to check that you have completed the following tasks:	
	Please mark "X"
Have you completed Sections A-E?	
Have you included copies of any documentation produced in support of your application? ***If the appropriate supporting documentation is not included with your application, you will then be asked to provide this separately. This may well delay the ethical review process, and thus the start of your research.***	
Have you signed as principal researcher and gathered signatures of the student (for student research) and department nominees as appropriate?	
Have you declared conflicts of interest (if any) in Section B qu 14?	
Are all pages (including supporting document attachments) numbered?	
Have you defined all technical terms and abbreviations used?	

How to submit the completed checklist

- ***If your department has a Departmental Research Ethics Committee (DREC), submit this checklist and supporting information to the appropriate departmental officer. To see a list of DRECs, go to*** www.admin.ox.ac.uk/curec/about/sshidrec/drecs/

- ***If your department does not have a DREC, submit the checklist and supporting information to the SSH IDREC (email ethics@socsci.ox.ac.uk). NB applications must be sent by email. Please do not send applications by post.***

Note on how to send signatures needed in Section E

- ***If you have obtained handwritten (wet-ink) signatures, please scan them plus the rest of the checklist pages to create a single PDF document and email through.***

- ***Alternatively, we accept "electronic signatures" sent as email confirmations from a University of Oxford email address. Separate emails should come from each of the relevant signatories as outlined in Section E, indicating acceptance of the relevant responsibilities.***

IDRECs and/or CUREC will review a sample of completed checklists and may ask for further details of any project.

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Director Professor Ernesto Macaro
Contact Natalie Usher, DPhil student – natalie.usher@education.ox.ac.uk

February 2015

Participant information sheet

Dear Student,

Study title: Learning about academic writing through peer review

In Trinity Term 2015, I am piloting a series of exam writing workshops for first year English and Joint Honours students. The aim of the workshops is for students to learn about and develop exam writing through peer reviewing a range of example essays, writing a timed essay and receiving feedback from peers. The focus will be on Prelims Papers 3 and 4 (Literature in English 1830-1910 and Literature in English 1910 – present day). The workshops will be evaluated as part of my doctoral research.

You are invited to participate in the workshops and the associated study. This document gives information about the study, explains what is involved, your rights as a participant and who to contact if you have any queries. The attached consent form specifies the conditions under which voluntary informed consent is offered and asks for your signature to indicate consent should you agree to participate in this research.

Background and aims
The aim of this study is to investigate what and how students can learn about writing from peer assessment. Prior research indicates that evaluating others' work is beneficial to students' writing. However, studies haven't looked at how this happens in much detail. So, the key aim of this study is to look at the impact of evaluating writing on students' writing processes as well as the quality of finished essays. Traditionally, writing is not explicitly taught as part of degree programmes. This study should give valuable insights into how peer assessment could be integrated into courses to support students' writing development.

I, Natalie Usher, am the main researcher. I am a doctoral student at the Department of Education and Wolfson College, and my professional background is in English teaching. I am supervised by Dr Therese Hopfenbeck, Lecturer in Educational Assessment at the Department of Education and Dr Kathleen Quinlan, Head of the Educational Development Group at the Oxford Learning Institute. We may also be assisted by a transcriber and a doctoral student providing reciprocal research assistance, who will be asked to sign confidentiality agreements.

Why have I been invited to take part?
You have been invited to take part in the study because you are a first-year student taking English Language and Literature or Joint Honours at the University of Oxford.

Do I have to take part?
It is completely up to you whether you decide to take part in the study. It must be emphasised that the exam writing workshop will be offered in addition to the teaching you normally receive. It will not affect your normal teaching in any way.

Before you decide to participate, please ask any questions you have. If you choose to participate, it is your right to withdraw yourself and any data from the study without penalty at any time. You can do this by advising the researchers of your decision. There is absolutely no academic penalty if you don't want to participate or if you withdraw from the study.

What will happen in the study?
If you are happy to participate in the study, you will be asked to read and sign a consent form. You will be asked to attend the workshops and participate in some activities to evaluate them, as detailed below.

The workshops
If you decide to participate in this study, you will be asked to commit to attending a series of four workshops. It is important that you attend all four workshops. These sessions will last 1.5 -2 hours and take place in Weeks 1-3 of Trinity Term 2015. The workshop activities will involve reviewing and discussing authentic examples of exam writing. You will also be asked to write a timed essay on a question of your choice. This will be typed up and your name removed. You will give written feedback on three essays from another group, and receive feedback from three peers from the other group. You will also be asked to write a reflective diary entry at the end of each session.

Evaluating the workshop
In order to evaluate the workshop, I will need to collect a range of data. Firstly, I will take a copy of the work you produce during the workshops, and also ask you to provide samples of essays, e.g. collections essays and practice exam essays written before and after the workshop.

Secondly, during the first and final workshop, I will ask you to complete two short pen-and-paper tasks in which you read and evaluate a sample essay.

Thirdly, I will ask you to complete a short questionnaire in the last session with your perspectives on the usefulness of the workshops.

Finally, I will ask you to take part in a think-aloud writing activity and interview in weeks 6, 7 or 8 of Hilary term. Time permitting, you may also be asked to take part in this activity again in week 6 or 7 of Trinity term. In the activity, you will be given an exam paper and asked to write an exam essay on a question of your choice. Up until the point where you start writing the essay in full sentences, you will be asked to think aloud or voice your thoughts. For example, you will be asked to say what you are thinking as you read through the paper, choose a question and think about how you will address it. However, you won't need to explain what you are thinking. When you start writing, you won't need to think aloud. After you have finished the essay, you will be asked some interview questions about the process of writing the essay, your thoughts on the finished product, and about writing more generally. The think-aloud activity and interview will be audio-recorded and transcribed. I will also take in the essay, but send you a photocopy of it afterwards. Some people also find it interesting to read the think-aloud transcription, and I can provide you with a copy of this after the workshop is finished, if you would like.

This activity will last for up to two hours and take place in a quiet room at the Education department. In recognition of your time and contribution, you will be paid £20 in cash on each occasion you take part in a think-aloud writing activity and interview. You will be asked to sign a receipt, and the information on this receipt will be used for the sole purpose of processing the reimbursement of the payment.

Are there any potential risks in taking part?
We do not foresee any risks involved in participating in this research. Occasionally, people don't like the feedback they receive or find the way it is written upsetting. The workshop leader will foster a supportive atmosphere and students will be given guidance in giving thoughtful and constructive feedback.

We appreciate that you have heavy workload, and so the workshops are designed to fit in with revision and exam preparation activities. With regard to the think-aloud activity, you will be paid in recognition of your time. Should you take part in a follow-up think-aloud after the workshop, you will also be paid for that time.

What happens to the research data provided?
Your identity, any personal information, and the identities of people mentioned in any interview or email communication will be confidential. This means that you may choose a pseudonym, or we will assign one for you, so that your name is not directly associated with particular pieces of data. As soon as possible

after you provide handwritten documents, e.g. workshop activities, these will be typed up and only your pseudonym will be used in the digital file. You have the right to review all of your data at any time. At the end of the project the data will be stored and kept secure for a minimum period of three years after publication.

The only data that will be seen in its full and anonymised form by people outside the research team will be the essays you provide, that is those written before, during and after the workshop. The typed up, anonymised versions of these will be read and rated by tutors at a later date in order to gauge the impact of the workshop on writing development. The essays will not be used for any other purpose without your explicit consent.

All data will be stored in accordance with the Data Protection Act, 1998 and the University of Oxford guidelines on Research Data Management (http://researchdata.ox.ac.uk/) to protect participants against possible misuse of information about them held by us. For the purposes of the Data Protection Act 1998, the University of Oxford is the data controller. Specifically, all signed consent forms and hard copies of data will be kept in a locked filing cabinet. The digital data will be password-protected. Backups will be made via the university's HFS system, and on one external hard drive, which will be kept in a locked filing cabinet.

Will the research be published?
This study is being conducted as part of my doctoral research, and so will be written up as a thesis. With your permission, I would like to use direct quotes from you when writing up the research. In the write-up of the research, pseudonyms will be used. The university may be identified, but your college will not.

As well as my doctoral thesis, the research will also be written up for publication in peer-reviewed journals and for presentation at conferences. I also intend to share and present the results to the English faculty so that they can decide whether to adopt this workshop in the future. I will be happy to share the results with you, and if you would like, I will send you a summary of the results once the study is finished.

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

If you agree to participate in this project, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be published with open access, meaning available to every internet user.

Who has reviewed this project?
This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. The research design has been reviewed by my supervisors, and by two other academics as part of the doctoral student development process.

Who do I contact if I have a concern about the study or I wish to complain?
If you have a concern about any aspect of this project, please speak to Natalie Usher (natalie.usher@education.ox.ac.uk; 01865 284099) or her supervisors, Dr Therese Hopfenbeck (therese.hopfenbeck@education.ox.ac.uk; 01865 274017) and Dr Kathleen Quinlan (kathleen.quinlan@learning.ox.ac.uk; 01865 286824) , who will do their best to answer your query. The researcher should acknowledge your concern within 10 working days and give you an indication of how she intends to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the chair of the Research Ethics Committee at the University of Oxford (Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD). The chair will seek to resolve the matter in a reasonably expeditious manner.

In order for you to participate, the University requires that you understand the nature of the study in which you have agreed to participate. After reading this document, please complete and sign the attached consent form if you agree to participate.

Thank you for considering participating in this research.

Yours faithfully,

Natalie Usher

UNIVERSITY OF OXFORD

Consent form for *Learning about academic writing through peer review*

Researcher: Natalie Usher, DPhil student, Department of Education and Wolfson College
Supervisors: Dr Therese Hopfenbeck, Lecturer in Educational Assessment at the Department of Education and Dr Kathleen Quinlan, Head of the Educational Development Group at the Oxford Learning Institute

Purpose of the study: to investigate what and how students can learn about exam writing through peer assessment.

	Please initial the appropriate boxes
I have read and understood the project information sheet.	
I have been given the opportunity to ask questions about the project and have received satisfactory answers to questions and any additional details requested.	
I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.	
I understand that my taking part is voluntary; I can withdraw from the study at any time without academic penalty and I do not have to give any reasons for why I no longer want to take part.	
I understand that participating in the study is in addition to the teaching I normally receive and that it will not affect my status as a student.	
I understand that taking part in the project will include attending four workshops, taking part in workshop activities, taking part in at least one think-aloud activity and an interview, writing essays, completing a short questionnaire and providing samples of my writing.	
I understand that my words may be quoted and excerpts used by the project researchers in the thesis, publications, reports, presentations, and other research outputs but my name will not be used.	
I understand that the only data in its full form (i.e. from beginning to end) that will be seen by people outside of the research team will be the typed-up, anonymised version of the essays I write. These will be read and evaluated by tutors to gauge the impact of the workshop on writing quality. They will not be used for any other purpose.	
I consent to being audio-recorded for the think-aloud activities and interviews.	
I understand that audio-recordings will be neither heard by people outside the research project nor included in publications.	
I understand my personal details such as name and email address will not be revealed to people outside the project.	
I understand who will have access to my data, how the data will be stored and what will happen to the data at the end of the project.	

	Please initial the appropriate boxes
I understand who is the data controller for the purposes of the Data Protection Act 1998.	
I understand that, in recognition of my time, I will be offered a payment of £20 on each occasion that I participate in a think-aloud writing activity (that is, thinking aloud, writing an essay and participating in an interview).	
I understand that receiving payment for participating in the study does not affect my rights as a research participant.	
I understand that if I have a concern about any aspect of this project, I can speak to Natalie Usher (natalie.usher@education.ox.ac.uk; 01865 284099) or her supervisors, Dr Therese Hopfenbeck (therese.hopfenbeck@education.ox.ac.uk; 01865 274017) and Dr Kathleen Quinlan (kathleen.quinlan@learning.ox.ac.uk; 01865 286824) , who will do their best to answer my query. If I remain unhappy or wish to make a formal complaint, I should contact the chair of the Social Sciences & Humanities Inter-Divisional Research Ethics Committee, at the University of Oxford (Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD).	
I agree to take part in the study.	

Name of participant: ….....

Signature: …..... Date: ….....

Name of researcher: ….....

Signature: …..... Date: ….....

Appendix G: Ethically important moments

Ethically important moments are 'the difficult, often subtle, and usually unpredictable situations that arise in the practice of doing research' (Guillemin and Gillam, 2004: 262). The response to and reflexivity about these moments are important to research practice. This section discusses ethically important moments that arose around feedback, post-intervention writing sessions, informed consent, anonymity and emotion.

Feedback

Although at the beginning of the Workshop 3 the difference between feedback and a response was discussed, and ground rules were set, I was shocked by the tone of a small number of reviews in 2015. After transcribing the reviews, I highlighted sentences which were of concern, and discussed them with my supervisor, Therese Hopfenbeck. In most cases, despite the harsh tone, the sentences contained useful information. In three instances, reviews contained sentences (underlined), which came across as harsh without offering information. These are given below in Table G1 with commentary in the right-hand column. It was decided to redact the underlined sentences. This was a difficult decision as I wanted the feedback to be authentic, but I was very aware of the exam pressures and did not want the workshops to cause harm. The redactions were conservative as although harsh, the rest of the reviews contained information which the recipients could triangulate with other reviews.

In addition, at the beginning of Workshop 4 we discussed the importance of receiving feedback critically.

Table G1: Redactions made to feedback	
Amber to Tessa	
The introduction works well, though it took me an additional reading after reaching the end of the first body paragraph that you were arguing against the quote, not "for" it. There were also grammar issues that made it confusing. While you exhibit a great extent of control of the argument, knowing when to relate authors back to one another as well as what each author is contributing to the argument, there is no (or almost none) engagement with primary texts.	*In this sentence, the reviewer either seems to be comparing the essay to SparkNotes, or claiming that the essay is based on SparkNotes. Either way, the sentences does not add anything to the review – the impression of 'a lot of plot explication and surface analysis' is already given. This would be hurtful to an English student – SparkNotes*

Perhaps it is because three complex authors with a very large fictional outputs are being juggled, but there is a lot of plot explication and surface analysis. I asked myself after the first paragraph on Eliot if the details about those main characters & thrusts of plot couldn't have been quickly gleaned by a few SparkNotes entries, and the answer was no. The quotations from secondary sources were very good, so I doubt you're not engaging with the novels, but it's just not adequately made clear by the essay. Overall very good style – just perhaps drop one author or use primary texts more.	*is a low quality study guide website.*
Heather to Rachael	
– good detailed introduction, although argument needs to be stated more clearly and directly – in first topic sentence link idea of light to description – first paragraph lacks direction – is it light, sensuality, symbolism, subtexts? a topic sentence relating these to the question would help this – really good analysis of primary texts – good return to argument w reference to "subtexts" (paragraph 2) – conclusion introduces too many ideas and fails to reach a position – "the question remains": At the outset decide on an argument and argue it. it is up to you to answer the question in the course of your essay o however, good to raise issue with the quote & approach the ideas in the quote critically	*Rachael, the recipient of the feedback, had run out of time during the essay and in an interaction at the end of the workshop, seemed a bit down. The underlined sentences uses imperatives and appears to talk down to the recipient by stating the obvious: 'it is up to you to answer the question'. The feedback giver has attempted to soften the comment by adding a further praise comment in the sub-bullet point below, but given that I was concerned about the students' emotions, I decided that this comment could potentially be harmful.*
Heather to Ben	
– argument appears confusing in the introduction – ~~be direct and convincing by arguing something that makes sense~~ – the view of Derek Walcott is potentially saying something quite different and you do not give sufficient justification for the position you take – question asks what critical insights are gained by conceiving English literature as national and/or transnational, not what is problematic concerning about such a conception. ~~Maybe~~	*Again, the underlined sentences use imperatives, and the solution is not really a genuine solution – 'argue something that makes sense'. The reviewer has not taken into account that argument may make sense to the writer. The redacted phrase does not add anything to* *The second comment again uses an imperative and tells the writer to*

~~start by thinking through what the question is asking.~~ – pay attention to syntax – state clearly in introduction which texts/authors you will discuss to consolidate your position – could also improve arguments by discussing what a critical position seeing lit as transnational would mean – several points – <u>about African people as relics</u>, as <u>Plunkett epitomising a certain perspective</u> - are not properly explained and therefore risk subverting your argument – good knowledge of primary texts & critics – could be supplemented by historical contexts	*do something rather obvious. The reviewer has no insight into the writer's process and so this comes across as condescending.* *Ben's response to the feedback is discussed in Section 8.7.*

Informed consent as an ongoing concern

Further ethically important moments arose in 2015 in the post-intervention sessions. I was aware of students' busy schedules: participants were writing a portfolio in Weeks 4 and 5 of term and had Prelims examinations in Week 8, so I arranged meetings with participants at their convenience from Week 6 of term. All opted to come in Week 7 or even the beginning of Week 8. I emphasised to participants that I wanted this meeting to be as helpful as possible to participants' exam preparation. As some had found the think-aloud protocol difficult, I stressed that it was optional as I wanted the practice essay to be helpful given the proximity to exams. Ethically important moments arose in the case of two students who initially said they would do the think-aloud. I sensed some hesitancy and emphasised that they didn't have to if it wasn't helpful, and at this they decided to not do it. Consent was thus treated as an ongoing concern.

Informed consent again arose as an issue in the 2016 workshops. Following strong encouragement from their tutor, all of the year group from one college signed up to the workshops, so I tried to be particularly alert to informed consent issues. One ethically important moment arose with a student who had missed the second hour of Workshop 1 due to another commitment. She had been due to catch up with me prior to Workshop 2, but didn't come. During Workshop 2, I asked her if she'd like to catch up at the end, and she agreed and seemed enthusiastic. She apologised for not coming to catch up and said that her coursemates had given her a general summary of the discussion. We spent around half an hour discussing the example essays. It became apparent during the discussion that she didn't know the essays very well and hadn't made many annotations. At the end of the meeting, it

became apparent that the participant wasn't very happy with the essay she had written in Workshop 2. She had tried out a new planning technique which took half an hour and as a result hadn't finished the essay. She also seemed unaware that the essay would be peer assessed, despite this being made clear in the invitation email, study information letter, verbally, and on slides at the beginning of Workshops 1 and 2. I reassured her that it was just a practice essay and that the feedback was formative but sensed that not all was entirely well. This conversation came at the end of a three hour teaching session - I had caught up with another student before the 2-hour workshop began - and I recognised later that evening that I was thinking more like a teacher than a researcher and that it was unclear whether the participant had actually given her informed consent for the study. I discussed the situation with my supervisor (TNH) and we decided I would send the participant an email. This expressed my concerns about informed consent, and asked whether she was happy to continue. If not, I offered that she could spend no more than half an hour working on her essay and email it back so I could include it in the study. I also reminded her that she was free to withdraw from the study at any time.

The student replied that she'd like to work on the essay for a bit longer and duly returned it. In the workshop, the participant had written 304 words. The essay she returned was substantially longer at 1358 words and had been significantly revised as well as added to. It seems unlikely that this had been completed in the half hour I suggested. The essay was included in the peer assessment in its full completed state. As the mean word count for essays in the study's collection is around 900 words, I was concerned that including it could have skewed participants' perspectives of the ideal length of exam essays and made writers concerned about the length of their own work. However, the reviewers did not praise the length of the essay, and all three picked up that the essay became less relevant or looser as it went on, which suggests that they were able to contextualise the essay with others they had read, and not take length at face value. Post-feedback reflections do not suggest that the essay length was a concern. In the Workshop 4, I gave the participants word counts with their essays and I again reiterated that the average of the essays in the study was around 900 words. We discussed that it wasn't absolutely necessary - or physically sustainable - to handwrite three very long essays in three hours.

Anonymity

A further concern that arose in 2016 was that of anonymity. While in 2015, essays were reviewed by peers from other groups and so other colleges, in 2016 an imbalance between the groups meant that participants in Group 1, the intact college year group, read at least one essay and received at least one review from a students from their college. I was concerned about this as it opened by the possibility of participants being able to identify the writers and reviewers. I informed participants of this at the beginning of Workshop 2 and asked whether they were okay with it and also told them that they could email me if there was a person they did not want to review or be reviewed by. A question was asked about how many reviews they would get and whether they would be identified in the process, and I explained that all essays and reviews would be typed up and only referred to by number. The group consensus was that it was okay. When students received their feedback, I included a reflection sheet to gauge the impact of some reviewing within the group. No one raised concerns.

Appendix H: Extract from analytical memo on Megan's large-scale adaptation

Four key areas of adaptation are identified in the following tables:

1. Increased focus on question – using keywords (W2), revisiting question both through monitoring and with topic sentences (Int 2)
2. Experimentation with planning
3. Making sure she finishes at Int 2 – more focussed intro and different structure (2 paras + synthesis)
4. Making time to re-read and revise essay (see revision to introduction; also link to self-efficacy).

Table H1: Finding intended adaptations in interview and workshop data

This table collates adaptations that Megan would do next time she wrote an exam essay, tracing whether she did indeed adapt as intended, and links to workshop learning. In the SRL coding guide, these are adaptation – future (SRL 6).

	Intended adaptation	**Did it happen? Self-report**	**Did it happen? Other evidence**	**Workshop link?**
Int 1	Um… yeah, I think I'd – I think I'd like to have a go at planning differently, cos I think my structure was a bit too like… blocked out, as it were? Um and so… um maybe I would have like tried to maybe include like the Barthes earlier and have that maybe as like a running… thread all the way through. If I could, I think.	Tries a different structure, but: W2: 'I wish I'd spent a tad longer planning to completely solidify my line of arguement' W3: 'Plan for a solid 15 minutes to set out a very clear line of argument.'	Word count for plan at Essay 1 increases at Essay 2 very slightly. Plan at Essay 3	N/A
W2	Yes - really drum out my 'compensation' arguement more sharply	N/A – task-specific	N/A – task-specific	N/A – task specific
W3	Plan for a solid 15 minutes to set out a very clear line of argument.	Adapts style of planning at Interview 2, unclear whether time planning different.	Textual evidence for changes in planning	Planning is a long-term concern

	Intended adaptation	**Did it happen? Self-report**	**Did it happen? Other evidence**	**Workshop link?**
	Make sure I think I am writing whilst I am writing rather than just thinking, i.e. be aware there will be a reader.	Still intends to spend more time planning after Interview 2 Reader: At Interview 1: 'I've never thought of that before. I don't think I was to be honest. I think maybe I was only thinking of [pause] like myself as the person writing it. Umm.. yeah, no I don't think I was [laughs]' At Interview 2, is thinking of her tutor.	Reader: N/A	Reader: 38 references to clarity
W4	> Topic sentences > having more quotes (quotes that are a little more relevant to my points) > A tiny bit more time planning to really solidify answer.	Topic sentences: Interview 2– 'leading in and out of paragraphs' (see below) Planning: see above	Textual evidence: see below *Quotes* not coded as a standard at Interview 2. Perhaps they are taken for granted? In both W2 and Int2 essays, uses same number of quotes from texts and critics: 8. No think-aloud, so cannot compare time for planning.	Topic sentences: Feedback 13 references to *quotes* in workshop data.
Int 2	I think I would just spend like a couple more, like a minute or two more planning, just so that [pause] that's really like concrete in my thoughts before writing.	N/A – beyond the scope of the study	N/A	N/A

Table H2: Adaptations Megan reports making in interviews and workshop data

This table collates adaptations or things that Megan reports doing differently in the particular essay, and also gathers evidence for whether this adaptation can be seen in other parts of the data. In the SRL coding guide, these are adaptation – new today (SRL 4).

	Self-reported adaptation	**Evidence**
Int 1	Um I think I planned for longer than I did in my collections because um yeah, basically [Laughs] because my tutor said that I should spend more time planning [Laughs] so I tried to do that.	N/A – beyond scope of study
W2	Yes I tried to explicitly reference the quotation. I think it made my essay more focus [sic] but perhaps demanded by engaged slightly better with modernist texts.	Textual evidence: see introduction below Workshop evidence: 73 references to *question* in workshop data. including 11 additions made during discussions. W1 most important criterion: 'Engagement with the question / quote – and picking up on such nuances'
Int 2	Umm yes. Umm... I like usually… Like, in the past I've like come up with like a rigid structure umm but I think… I don't know, just one of our tutors suggested maybe not doing that and just like jotting thoughts of where your essay is going to go rather than like trying to like be like 'I'll bring in this then, I'll bring in this then' because like things can shift. So I tried to do it a bit more like that. NU: OK. So it was more of a loose plan, rather than like 'paragraph 1 is this, paragraph 2 is that' Megan: Yeah, yeah exactly	Textual evidence: Planning – from structure to annotations. See below. Planning is a preoccupation since Interview 1, and tutor feedback on collections.
Int 2	I think I spent longer... like on the question itself, if that makes sense.	Textual evidence: annotations around the question (see below) Workshop evidence on question: see above Deconstructing the question: strategy from tutor

	Self-reported adaptation	**Evidence**
Int 2	Yeah, I think I've been consciously trying to make sure that I like lead in and out of paragraphs well	Interview evidence: acting on feedback suggestion. Not completely new to her, but confirms and gives a way of dealing with it. At W4, makes four references to signposting, and states intention to work on this. Interview data on feedback: 'Yeah, no they were like fair and helpful, actually. I think… umm… kind of [pause] not even like [pause] kind of making sure that I'm kind of leading in and out of paragraphs like well, was definitely a good thing for me to work on.' Reviewer gave suggestion: 'That was something that I was already aware of, umm but like [pause] err, I don't know, the feedback was quite good as it kind of like [pause] showed me what else I could do, or like what kind of, maybe like… buzzwords probably isn't the right word but kind of theme it could be leading into a sentence or… NU: OK. So it's sort of… so it's like you'd already sort of realised that that was something you could work on, and it helped because it confirmed that? Megan: It confirmed that, and… kind of like suggested, I think they suggested a way to… yeah... to deal with it or improve it…' She is aware of the problem with her text, as stated in W2 reflections: 'Also I jump a bit between texts. I wish I'd made the links more explicit as to why I was jumping to Regeneration' Textual evidence: refers back to question throughout essay; similar approach to W2, with keyboard. But also refers to the quotation twice after the introduction, and linking between authors:

	Self-reported adaptation	**Evidence**
		e.g. end of first main body paragraph: 'Coetzee approaches literary tradition as problematic for the coloniser as sympathy with the colonised will always be blocked by the existence within a different discourse.' Starts second main body paragraph: 'Walcott deals with literary traditions very differently, instead of a concern with the divide which exists between discourses,
Int 2	I kind want to make sure that I've been producing like focused intros umm moving about like my time management.	Workshops: Makes 14 annotations on example essays, including 6 additions made during discussion. Comments on introductions of all three peer essays. No mention in W3 and W4 reflections. Finishing in time seems to be the underlying goal here – a 'focused' intro is the standard that will help her meet it Textual evidence: see below. Also makes revision to introduction at end of essay.
Int 2	Change in structure: So… I don't know, like with this like I – I think I… attempted to like write.. like three paragraphs but I think… maybe I've learnt that maybe that two or two and like a… like a small sort of synthesis of like a kind of theme that is been running through it is better than trying to do three massive things.	Continuation from 'focused intros' above – again, finishing in time is the underlying goal. Textual evidence: Change in paragraph structure from essay 2 to essay 3 Workshop data: Awareness of time pressure in interviews, but not also coded as emotion. Comments on peer essays about running out of time. All three reviewers comment on the need for a more developed conclusion.

	Self-reported adaptation	**Evidence**
		No coding references to structure.
Int 2	Umm, I kind of only skim read err... the other bits, but I think so I have started, I don't know how this has gone in my collection, but I've also started giving myself a couple of minutes at the end just to go over it, actually, umm, which is something that I'd like never... done before. But I think also it strangely a confidence thing, as well, to have like the confidence, because everyone else is writing to the end and to have the confidence to be like 'this is my essay, I'm now just now going to amend it', is kind of a difficult thing sometimes. Um… So like I've been trying to do that because I think especially like... for my writing where I probably don't…. like obviously you're writing quickly but I don't like punctuate well enough the first time around, and I don't know, umm… occasionally you read things back and you realise that you've missed a word or you've… And I think I do that quite frequently, so I don't know, I think it's a good thing for me to do.	Interview evidence: refers to change in introduction Textual evidence: Edit made to introduction with asterisk * - additional sentence in brackets Workshops: This adaptation comes up in response to interview Q on strategy handout 31 references to argument and 38 to clarity; 2 references to technical accuracy. 7 references to technical accuracy.

Table H3: Tracing adaptations in Introductions

This table examines the textual evidence for Megan's adaptations in more detail.

	Interview 1	**Workshop 2**	**Interview 2**
Essay word count	969	924	917
Finished?	Yes	No – last para is a fragment. On time: 'Badly this time. I didn't get to properly	She finished her essay with time to check over it, although in response to the

	Interview 1	**Workshop 2**	**Interview 2**
		write a conclusion as I ran out of time. So no I could not read over / revise'	question about ten minutes more, she would 'solidify the conclusion'. One of the reasons she mentions for judging her June essay to be higher quality than her workshop essay is that 'it has a conclusion'. Finishing is partly figuring out what she can do in the time – two paras plus a synthesis?
Written plan word count	91	109	42 + 16 words of annotations around Q = 58 words
Written plan	Detailed, para by para	Detailed, para by para. But this time with reminders: 'remember about compensation' 'turn the focus directly to pan & war poetry' 'remember apostrophe's' [sic] Additional: quotation section with ideas	Annotations around the question Three bullet points Key word from question, underlined Points for each author she writes about
Introduction word count	151	113	109
Question	'Literature is no-one's private ground; literature is common ground' (Virginia Woolf). Discuss.	'Take the word "doubt": it gives us hardly any sensuous suggestion of hesitancy, of the necessity for choice, of static irresolution […]. Mr. Joyce recognises how	'His language, so familiar and so foreign, will always be for me an acquired speech.' (JAMES JOYCE, *A Portrait of the Artist as a Young Man*). How have postcolonial

	Interview 1	**Workshop 2**	**Interview 2**
		inadequate "doubt" is to express a state of extreme uncertainty, and replaces it by "in twosome twiminds" […]. This writing that you find so obscure is a quintessential extraction of language and painting and gesture' (SAMUEL BECKETT on *Finnegans Wake*). EITHER (a) How does modern literature compensate for the 'inadequacy' of language? OR (b) Discuss obscurity in modern literature.	writers approached dominant literary AND/OR linguistic traditions?
Introduction moves	Brief reference to claim in quote → Eliot as contrast → reference to meaning and understanding – sort of linked to quote → contrast with WW1 poets – private → modernist drama as rejecting common ground	Overview statement quoting the question → summarising quotation → war and compensating for inadequacy of lang	Theory → links to novel and to quote → clarifies paradox → again links with Coetzee → sets up contrast with Walcott *Bracketed text after asterisk added later
Introduction text	Despite the title-based claim that Modernism was a literary school of thought that was somewhat revolutionary and rejected tradition, the poems of T. S. Eliot, who is probably the biggest name in modernism, heavily rely on literary tradition. This does not make these poems inaccessible, however it does close of meaning for those who do not know much of 'the dead poets' that Eliot's work is so endebted to. Alongside this are the	Modern literature often compensates for the 'inadequacy' of language by drawing attention to its own failure to articulate a meaning precisely. In the quotation above Beckett asserts that the obscurity of writers such as James Joyce is what is so quintessential to their 'extraction of language', however this is not necessarily the case. I wish to engage with writing about war, much of which was happening around the time that Joyce	Derrida argues in Racism's Last Word that 'racism exists in language' and this is a concept that both Coetzee in his novel Waiting for the Barbarians and Joyce in his quotation engage with. Joyce explores the paradox existing within Post-colonial writing*, (that the Post-colonial is inevitably bound up within the discourse of the coloniser or the colonists: that it is 'so familiar and so foreign') and this is a concept which Coetzee certainly is

	Interview 1	**Workshop 2**	**Interview 2**
	soldier poets of the First World War who had very specific war experiences and this meant their writing, although not closed off, was somewhat private, in the sense that their 'experiences' were very personal. Many also believed that civilians could not write about war. In addition to this new wave modernist drama can be seen to reject a common ground as it moves away from and questions what would be culturally perceived as the common ground of theatre.	was writing, to convey the idea that writing that would not be considered obscure in a modernist sense was able to compensate for the inadequacy of language and did so by being aware of its own flaws as a piece of writing.	concerned with. Walcott, in comparison, does not allow there to be a dominant literary tradition or a dominant space of literature and allows the colonial and literary traditions to coexist in his epic poem Omeros.

Table H4: Standards

Query results generated in NVivo. For interview and TAP data, references are coded as a criterion code AND as standard are cross-tabulated by time. As the workshop data is less rich, references are coded as criteria only. The NVivo report was compiled for two purposes: 1. To link adaptations to workshop learning; 2. To identify any changes in standards across the study.

Green highlighting denotes top 10 codes.

- The following stay stable, i.e. in the top 10 across the three: argument, critics and theory, development, introduction, question
- The following increase at workshops and at interview 2: clarity, conclusion
- The following makes it into the top 10 only at interview 2: technical accuracy
- The following are only top 10 for workshop data: evidence, signposting

	Interview and TAP 1	**Workshop data**	**Interview 2**
C1 Analysis	0	5	0
C2 Argument	2	31	2
C3 Choice of authors or texts	0	4	1
C4 Clarity	0	38	4
C5 Comparison and linking	0	8	0
C6 Conclusion	0	15	3
C7 Consistency; coherence	1	6	2
C8 Context	0	3	0
C9 Critics and theory	4	14	3
C10 Depth	0	0	0
C11 Development	2	13	2
C12 Evidence	0	14	0
C13 Finishing	0	0	0
C14 General; broad	0	0	0
C15 Introduction	3	27	3
C16 Knowledge…	1	11	0
C17 Nuance; complicating	0	5	0
C18 Original	0	1	1
C19 Plot	0	2	0
C20 Progresses; direction	0	3	0
C21 Question	5	73	10
C22 Quotes	5	13	0
C23 Range; scope	3	5	0
C24 Relevance	0	3	0
C25 Signposting…	0	24	1
C26 Specific, selective	0	0	1
C27 Structure	7	1	2
C28 Style	0	4	0
C29 Style – concise, precise	1	3	0
C30 Technical - referencing	0	1	0
C31 Technical accuracy	0	7	2
C32 Terminology	0	0	0
C33 Voice, positioning…	0	9	0
C34 Word count; length	1	1	1

Appendix I
Coding guide: criteria

The guiding question for this coding is: *What* is the participant commenting on in this response / annotation?

Examples are given for each type of data for which the code is used, and in the following order: workshop data, interview data, think-aloud protocol (TAP) data (as applicable), feedback data. ***Bold and italics*** used to designate other codes – e.g: ***comparison and linking***.

Code name	**Explanation**
C1. Analysis	Exclusively used for specific reference to 'analysis' or forms of the term. This often seems to mean analysis of quotations. 'Interesting analysis but perhaps needs an explicit link to the question.' (Celine) 'Whilst some analysis is really interesting, particularly in the first paragraph, the majority is plot-based statements that hamper rather than develop the argument.' (Chloe) 'There is lots of nice, close analysis of quotations used, but no critical quotations or awareness.' (Joe) 'I tried to do a fair amount of… quoting the primary texts and like close analysis of that? Umm yeah because I quite focus on… devices, so I… I did talk a lot about metaphors and umm kind of what makes a description vivid, or what makes it realistic.' (Nicole, Interview 2) 'I suppose, normally I would say analysis [is a strength], but because I didn't really know any quotes there wasn't really any particularly close analysis.' (Heather, Interview 1) 'None of your essay simply repeats plot points – instead you are really good at analysing precise instances within the texts in relation to critical ideas and significations.' (Gemma, giving feedback) 'It would have been nice to see some instances of close analysis' (Ben, giving feedback) Overlapping with other codes allowed. As seen in the example above, relates closely to ***quotes*** – quotations allow participants to do

Code name	Explanation
	'close' analysis. Also relates to ***depth*** – 'And also I could have gone a bit deeper in analysing it' (Felicia, Interview 1).
C2. Argument	Broad code used for the term 'argument' or forms thereof; also thesis. 'This essay does not seem to have left itself time to develop an argument, let alone one which would address the terms of the question and the complexity of the quote which was given.' (Abbie) 'The writer did well to explain their thesis in the introductory paragraph, making sure to explain the way they wanted to view and set up the complex terms of the question.' (Zoe) 'Um, so after the plan, I put - I started straight away with my introduction. I did make a false start because I was just, I was getting used to it and I thought, "No, that sounds really awkward." And I took some time and I said, "OK, what do I really want to say?" and then started again and thought, "OK, this is my argument."' (Holly, Interview 1) 'Umm I think I'd want to like clarify my argument [pause] at the beginning, because reading it back, like I think the intro's OK, but it could be stronger, and it could be more like directly focused then.' (Megan, Interview 2) 'Umm again, that's not quite looking at what my interests in are, which is religion and science and faith and action is purely focused on Tennyson and I probably wouldn't have a strong argument there, I'd be able to just write about it.' (Phoenix, TAP 1) 'There are also just a few points where the links between textual evidence and argument are more implicit than explicit and could perhaps just to be clarified is to strengthen the argument (e.g. some of the points you make about Clough and liberation).' (Celine, giving feedback) 'While you exhibit a great extent of control of the argument, knowing when to relate authors back to one another as well as what each author is contributing to the argument…' (Amber, giving feedback)

Code name	Explanation
	Relates to other criteria – overlapping allowed.
C3. Choice of authors or texts	Comments on the writer's combination of authors or texts – whether the reviewer felt the author made a relevant or interesting selection of texts. More global than ***comparison and linking*** – comments focus on overall impression, or focus on the need to justify or frame the choice of texts, or to acknowledge the underlying differences between texts (e.g. time period, literary movement), usually in the introduction. These latter comments tend to be made as additions. Less frequent in interview data. 'Good parts of this essay is that it ties together texts that might not otherwise be grouped in a unique fashion' (Amber) 'Combining well known and lesser authors.' (Abbie) 'Four v diff novels treated as exemplar desp diff not addressed → acknowledge differences.' (Heather) 'Good that they've used Woolf, as she is quoted in the question.' (Rachael) 'These are all really different texts. Good to acknowledge differences + similarities in the texts.' (Zoe) 'Umm and the texts kind of fit well, but they're also kind of like an unusual combination to kind of to put together umm and so like bounce off each other kind of nicely.' (Megan, Interview 2) 'Your sense of literary history and chronology of your broad choice of texts is also made clear and is a very strong element of your essay-writing style.' (Claire, giving feedback) '…but what's unclear is how these three writers are truly "fruitful" points of comparison' (Amber, giving feedback – i.e. peer needs to justify choice of writers) Similar to ***comparison and linking***, which includes references to *how* writers use the texts they choose – ***choice of authors or texts*** operates at the more macro level. References cannot be coded at both.
C4. Clarity	References to the reader's impressions of how clear or unclear the essay is to read. In interviews, includes the writer's impressions of their own essay, or their intentions to make their essay clear and easy to read. Key words include clear/unclear and synonyms – e.g. explicit, confusing. In example essays, includes instances highlighting how the essay was written clearly, flagging up the need for

Code name	Explanation
C4. Clarity (cont'd)	more clarity or explicitness, or commenting that something becomes 'muddled'. 'Successful: the student writes clearly and authoritatively, employing critics at an appropriate frequency throughout.' (Sean) 'Not clear what they are discussing – prose? plot? Seems to do both but in a lack of depth.' (Gemma) 'No idea what this sentence means.' (Amber) 'Um… but also… and I think as a result of that my argument was less clear in this essay. I didn't really sit down and think, "Well, what exactly am I arguing?"' (Holly, Interview 1) 'I think structure is similarly important, I think like – it just gives you a direction and to the reader of your essay it just makes it really clear.' (Chloe, Interview 2) 'I had to make it clear when I was, like yeah, when I was saying about straying away from talking about description, I tried to make it clear my essay, OK well this isn't exactly description, this is another thing that works alongside description, so just trying to qualify everything that I was doing.' (Nicole, Interview 2) 'HOWEVER the author does not explicitly state what CRITICAL INSIGHTS are gained, as the question asks. They explore it implicitly' (Jess, giving feedback). 'They seemed to fuse together which created a sense of confusion and lack of focus on the question.' (Rachael, giving feedback) 'At times points lacked clarity of expression' (Poppy, giving feedback) In example essays, ***clarity*** also links to readerly perspective. This example implies Heather interpreted the quotation in the question differently and that she considers the author's understanding to be flawed. 'However would be more clear, especially in intro if quote better understood.' (Heather) May overlap with other codes, which often provide explanation for the reader's impression of the (lack of) clarity, or suggestions for

Code name	Explanation
C4. Clarity (cont'd)	how to resolve this. The following example is also coded as ***consistency and coherence.*** 'Perhaps at times the focus just becomes slightly muddled in exploring the ideas of time and home, which do link well to the argument, but perhaps this link just needs to be articulated more explicitly/clearly?' (Celine) 'focus… muddled': ***clarity*** – reader's impression. 'this link just needs to be articulated more explicitly/clearly': ***consistency and coherence*** – suggests how the lack of clarity could be remedied. In the following example, where Felicia is talking about what she would do if she had ten more minutes to work on the essay, ***signposting*** is also used as a code as this is a strategy she could use to make her intentioned point or argument clearer: 'Uh, read back over. And maybe, like towards the end of paragraphs, add in like another sentence just to really summarise and make explicit what I was trying to say in the paragraph. Because I think sometimes I got a bit confused towards the end.' (Felicia, Interview 2)
C5. Comparison and linking	Refers to how the writer deals with the texts they write about – the presence/absence of comparison and/or linking to other authors. An example from the feedback data: 'Perhaps your argument could have been made clearer and sharper by the inclusion of clear paragraph-openers to flag up where next your essay is going, and through the inclusion of a conclusion to tie your main points together at the end.' (Zoe, giving feedback). Can refer to the structure of the essay and how writers use texts together. 'The running comparison with the works of writers really effective, as he functions as context, critic and a further literary material. The essay benefits from the comparison with 'The Return of the Soldier', which helps to emphasise the war aspect of the answer.' (Joe) 'I like that the texts are discussed through points + then similarity + difference, rather than first one and then the other' (Celine) 'A comparison here, but elsewhere not much comparison' (Sean)

Code name	Explanation
C5. Comparison and linking (cont'd)	'This essay is very clear. The comparisons made between Woolf and West were skilfully done; the two authors have been integrated well into the argument' (Zoe) 'Umm… well, I think comment on Clough's quotation, you have to kind of assess the truth value of it, and then if they're asking you to talk about more than one poet, surely they're looking for a comparison or a contrast in relation to the quotation.' (Amber, Interview 2) 'Erm, [pause] I like the way at the end I do compare it to – I compare both together and then both of them with social context' (Ben, Interview 1) 'Umm. I could compare it a bit to Elizabeth Gaskell, if I could remember a quote.' (Ben, TAP 1) 'But the deeper comparison between the authors as well as wider social significance that you go into in the conclusion looks promising.' (Nicole, giving feedback) 'The writers were really nicely combined and returned to' (Abbie, giving feedback) '…sometimes the links between texts and paragraphs were tenuous' (Joe, giving feedback) Related to ***choice of texts or authors***, which focuses on which texts and authors the writers select, and how they frame or position their selection, e.g. in the introduction – i.e. at a more macro level. ***Comparison and linking*** refers to how writers use the texts they choose. References cannot be coded at both.
C6. Conclusion	Any mention of the conclusion – e.g. its clarity, the role it plays in the essay, For workshop data, this code includes annotations made specifically about the conclusion, or final paragraph: 'Nuanced conclusive comments, interpreting the quote and question in a sophisticated manner' (Claire) 'I felt the argument vacillated throughout □ end did not move to explain a precise judgement' (Poppy) 'The conclusion is very convincing and shows the student has a broad knowledge of imperialist/colonial discourse' (Sean)

Code name	Explanation
C6. Conclusion (cont'd)	'Err, I think I'd work on [laughing] on the conclusion a little bit because I think I've ended up just sort of summarising and repeating my ideas. Umm, my tutor always talks about the extending implications [laughs] so I think [laughs] that would have been a good idea' (Tessa, Interview 1) 'The conclusion is quite brief and could have been more developed and outward-looking.' (Rachael, giving feedback) 'Furthermore, the conclusion could have been improved by referring back to topic sentences which would have drawn out implications and further structured the essay.' (Holly, giving feedback) Can refer to instances where participants run out of time to write a conclusion: 'I mean I did end up shifting things around and changing things and I hadn't quite managed to write my kind of conclusion [laughs].' (Nicole, Interview 1)
C7. Consistency, coherence and focus	References to the extent to which the writer's argument or discussion is consistent and/or coherent – whether the argument is sustained, focussed. References in interview data often are to the writer's attempts to make the argument consistent throughout, or to use a thread throughout the essay. This code is also used for inconsistencies and contradictions; jumps and leaps. 'The introduction lacks an argumentative slant specifically, but it addressed the issue at hand well and succinctly and the essay as a whole was argued with coherence and attention to detail.' (Claire) 'Good but again a leap from paragraph before' (Ben) 'What I like: it stays on topic, though sometimes does not address the "capitalist" part of the question' (Sean) 'Perhaps at times the focus just becomes slightly muddled in exploring the ideas of time and home, which do link well to the argument, but perhaps this link just needs to be articulated more explicitly/clearly?' (Celine)

Code name	Explanation
C7. Consistency, coherence and focus (cont'd)	'I felt the argument vacillated throughout – end did not move to explain a precise judgement' (Poppy) 'I would develop my conclusion and read through back over it and try and make sure it is kind of coherent thread throughout; maybe like change a few things to make sure it fits altogether a bit better, rather than being a bit more fragmented' (Felicia, Interview 1) '[Laughing] I think it's because I get [laughs] halfway through it and then think "Wait, where am I?" [laughter] I've got all these points but I don't know where to go with it or you have another idea, umm, that you wrote at the beginning but you don't quite know how to link it to the idea you've just said. So you don't want to just insert a random, I don't know because you want your essay to be cohesive, and have sort of argument running through it.' (Tessa, Interview 1) 'So I think as it was, I maybe wrote about it bit too much. And then I think it almost comes across a bit incoherent, because it's so obvious that what I'm saying is that realism. You know, they're not these perfect women, but they make mistakes and everything. So it's obviously because it's realist literature, but I don't actually discuss [laughing] realist literature.' (Heather, Interview 2) 'I kind of looked back at the question just to to check that that was like the focus, and umm and then stopped again looked again at the end of the – no just before I did the conclusion, just to make sure that was kind of linking up.' (Megan, Interview 2) '….but contradictory to other parts of the essay ("Where Faulkner differs from Woolf")' (Rachael, giving feedback) 'With this essay you just have to run the argument through all the way to the end' (Megan, giving feedback) '…and is immediately justified by directly engaging with the terms of the quotation, <u>which it continues to do throughout</u>.' (Joe, giving feedback) This is different to ***development***, which refers to how far the writer pushes their point or argument. It is also related to but distinct from ***structure*** which is used for references on the macro structure of the essay. Readers may point to 'jumps' or 'leaps' or 'straying' between paragraphs or points. As with other references to ***consistency***, these are only coded as ***signposting*** as well only when it is apparent that the reader attributes these to lack of signposting, e.g.: 'The topic sentences at the beginning of each paragraph are clear, but the development of the argument tends to stray and lacks focus.'

Code name	Explanation
	(Tessa) In the case of a writer, ***signposting*** is used because the argument is complex and thus less intrinsically cohesive: 'I reckon the essay I've just written, the umm... Coetzee and err Derek Walcott essay, umm is theoretically deeper, but as a result is a little less cogent and cohesive. Umm and I've done my best to signpost the argument but I think because the argument is a bit more complex, I - and I haven't fully comprehended how I was relating it to the question it's a little bit more woolly, I'd say.' (Phoenix, Interview 2) No references in TAP data.
C8. Context and period	Any mention of context or contextual knowledge - both literary and historical. Also includes references to the period of literature, and to situating the writers and texts within the literary period, or attempts to deal with literature from across the period studied and assessed on the papers (1832-1910; 1910-present day). Initially, there was a separate code for *period* as this is a recurring term, but the codes were merged as there was much overlap with context. 'Good to use context but statistic confusing' (Jess) 'Like how they immediately ground the argument in context' (Felicia) 'Awareness of previous literary movements' (Claire) 'Progress of lit over period' (Gemma) 'Recognising in the scope of the period → perhaps back it up contextually with dates.' (Joe) 'Though I think… making the um [pause] link from the kind of looking quite closely at the words and the way that they function to making comments about [pause] a writer and what they're saying about their society or if they're saying anything at all, that kind of process I want to develop more.' (Nicole, Interview 1) 'It [the question] also says 'writing of the period' so it kind of needs to be across the period.' (Felicia, Interview 1) 'I like the way at the end I do compare it to – I compare both together and then both of them with social context. I think that's a good to [pause] not generalise but to get a bigger perspective of the literary context.' (Ben, Interview 1)

Code name	Explanation
C8. Context and period (cont'd)	'But I'd probably say [pause] tried to talk about a scope [pause] of modernism that goes from Mallarmé, which is like practically pre-modernism to… Beckett, which is practically post-modernism. I think that's probably…. ambition, there you go [laughs]' (Nicole, Interview 1) 'So whereas we were taught kind of topic by topic, but didn't necessarily relate, it's trying to make sure that I'm filling in the gaps, I have a good awareness of the whole literary period.' (Tessa, Interview 2, talking about revision). 'Paragraph 3, umm, context, question mark? [writing] Separate spheres, umm, marriage acts, and women's colleges, mmm, convention equal religion equal piety equals support identity of husband.' (Holly, TAP 1) 'It's got loads of contextual evidence.' (Rachael, giving feedback) '…and focused interaction with both text and context.' (Poppy, giving feedback)
C9. Critics and theory	Any reference to the use of critics / critical detail / critical ideas / critical themes / critical debates / theory / critical awareness. Also secondary sources, where these are not explicitly stated to be context. 'Good use of textual reference – perhaps need critics?' (Poppy) 'So many critics – too much?' (Abbie) 'Interesting that secondary criticism is added before the textual example' (Claire) 'The writer is evidently familiar with both texts, relevant literary theory and is possessed of sufficient flexibility to adapt their knowledge to meet the demands of the question.' (Holly) 'I think maybe like that it touches on the theory at the end, I was quite pleased with myself for getting that in and trying to link it back and questioning sort of literature from the viewpoint of an author as well as a reader' (Megan, Interview 1) 'I mean I tried to link it to some critical theory, not sure how successful it was, but I tried [laughs]' (Felicia, Interview 2)

Code name	Explanation
C9. Critics and theory (cont'd)	'Could include some critics, Bhabha, Said?' (Tessa, TAP 1) 'I also feel that the texts you have used have been well integrated into the essay, especially the literary criticism by Said.' (Zoe, giving feedback) 'The critical knowledge was well deployed throughout so too was knowledge of the texts.' (Washington, giving feedback) 'You could make this essay a lot stronger with the help of some critics.' (Sean, giving feedback)
C10. Depth and detail	Use of the term depth, or to attention to detail. Often co-occurs with ***analysis*** or ***knowledge***. 'In-depth knowledge of the whole text [s1: 'parenthetically' underlined]' (Joe) 'A good, detailed response with in-depth focus on two texts and a broad range of critical evidence and theory backing up the point.' (Phoenix) 'The introduction lacks an argumentative slant specifically, but it addressed the issue at hand well and succinctly and the essay as a whole was argued with coherence and attention to detail' (Claire) 'And also I could have gone a bit deeper in analysing it.' (Felicia, Interview 1) 'The problem with this I wasn't practiced, I hadn't, I had been a bit out of practice in writing under time conditions, so it's quite rushed. And it also doesn't get as far in terms of that higher in-depth' (Ben, Interview 1)' I kind of find myself in a Prelims essay like, maybe slipping into the level of detail that I would in a tutorial essay, and I have to kind of stop myself because I'm like, "your hand is not gonna move this fast". Um, so, you have to kind of… if your lens is really kind of focused-in you kinda have to go back to the bigger picture a bit' (Amber, Interview 1) 'Good detailed introduction' (Heather, giving feedback) 'Overall thorough, detailed and confident.' (Poppy, giving feedback) 'More depth in some earlier points & and less breadth here?' (Ben, giving feedback)

Code name	**Explanation**
C10. Depth and detail (cont'd)	Can overlap with ***range*** – participants talk about the need for doing both in a Prelims essay, or too wide a range of texts precluding depth. 'Also, you would expect it, even though you need to have range and depth, you would expect a bit less depth in a prelims essay realistically. Like, you might mention a critic, but you're not going to engage them as well as you would be expected to in a tute essay.' (Heather, Interview 2) 'However I felt the essay would have benefitted [sic] from less authors; Disgrace is only mentioned in passing and the range precludes any in depth analysis relevant to the question' (Holly) Similar to ***development***, but references cannot be coded at both. ***Depth*** relates more closely to level of detail, ***knowledge*** and ***analysis***; ***development*** relates more closely to arguments and points, and pushing them to fruition.
C11. Development	Any reference to development of a point or argument, or lack of thereof. Includes annotations where the reader exhorts the writer to say more, and feedback asking for more elaboration. Common synonyms: sustain, explore. Also explain? 'Each segment of the argument is well-developed in meaty paragraphs, whereas West is less heavily emphasised' (Amber) 'I feel that there are places where a new point is introduced (e.g. the link to 'Hard Times') but then abandoned without being fully developed.' (Celine) 'It feels well-developed in relation to Mary Barton but perhaps less so to the reception/criticism of the novel.' (Rachael) 'How? Why not?' (Abbie) 'More sustained' (Ben) 'Nice but keep going' (Sean) In interviews, can refer to adding:

Code name	Explanation
C11. Development (cont'd)	'Umm I would… add another paragraph, which would nuance my argument in the way that I wanted it to go, because I definitely had another paragraph I could have used, which would have kind of made it a bit more… complex and just bring up the level, really.' (Nicole, Interview 2) 'Um… I think [Pause] er I would try and make it a bit more [Pause] precise and kind of wish I'd like left a bit of space to maybe develop some of the points a bit further. [Pause] Um. [Pause] In terms of maybe like exploring what…. like exploring Eliot's actual quotes and maybe thinking about how… that relates a bit more back into his theory, if that makes sense' (Megan, Interview 1) Examples from feedback: 'At times I felt further elaboration was needed to improve the forcefulness of the argument (the last sentence of the 2nd para for instance).' (Washington, giving feedback) 'Saying this the argument was still developing later in the essay and the ultimate conclusion felt like a powerful one which perhaps just deserved some more space to unpack it?' (Abbie, giving feedback) 'The essay would benefit from a more extensive conclusion to tie up the many, diverse explorations in the essay.' (Joe, giving feedback) Some overlap with ***clarity*** – (lack of) development can impact upon this. References cannot also be coded at ***depth***. ***Depth*** relates more closely to level of detail, ***knowledge*** and ***analysis***; ***development*** relates more closely to arguments and points, and pushing them to fruition.
C12. Evidence	Any reference to the presence or absence of ***evidence***, examples or detail, most often giving support for a point or argument. Occasionally, ***evidence*** is of the writer's knowledge. 'Textual, contextual and critical detail is thorough and deployed in combination' (Abbie) 'Vague. No evidence' (Tessa) 'Maybe good to have textual example here' (Jess)

Code name	Explanation
C12. Evidence (cont'd)	'Makes a lot of generalised statements that are left without substantiation.' (Megan) 'Opening is grounded in an immediate contextual basis, supported by secondary material [s2]' (Joe) 'I am not sure why the student brings in biographical details about Carlisle [sic]' (Sean) 'Umm… I think it's quite well-informed [laughs]. I think I've got some good primary knowledge, which is something I really struggled with at the start of year, I had argument that wasn't really grounded in the texts. So what I've done is I've gone back and I've looked at my primary materials and gone, well where is there evidence? And I think that's probably a strength of this essay.' (Holly, Interview 2) 'I wouldn't say [laughing] it has a lot of evidence' (Chloe, Interview 1) 'Perhaps another example of non-standard language in this work & its effects.' (Jess, giving feedback) 'I think the evidence and writing is sound.' (Megan, giving feedback) 'Also, could you include a critic here who suggests that Eliot had an inability to write a better poem, as this will allow you to engage in critical arguments and would help avoid it seeming like an unfounded assertion.' (Gemma, giving feedback – asking for more evidence. Also coded as ***critics and theory***). Quotes are a key form of evidence, and there is a separate code for ***quotes*** – the term has to be used and 'textual evidence' and 'textual example' is coded as ***evidence*** as they are broader. No items should be coded as both.
C13. Finishing	Any reference to managing – or, more usually, not managing – to complete the essay in the time. Can include doing everything you intended to do; having a conclusion. This code first emerged from the SRL coding of interviews and TAPs. To ensure consistency, ***time and process*** (now under 'other codes') was refined to exclude any items which were focussed on finishing (or not) in time. A text query search on the workshop data for 'finish' and stemmed words only retrieved one further example which fits this code. I expect more instances to come up in

Code name	Explanation
C13. Finishing (cont'd)	the feedback data, as there are more 'unfinished' essays than in the example essay set. 'There is also no clear conclusion, the whole effect is that of an essay that has been rushed' (Chloe) 'Is it even a conclusion or just another paragraph which they ran out of time to finish' (Felicia) 'Because my last…I had a collection the beginning of term. I sort of ran out of time quite a lot so I was conscious of that and sort of getting things like a conclusion in rather than sacrificing other parts.' (Chloe, Interview 1) 'I was more focused on moving forward and trying to… trying to get to the end. There were bits… erm I definitely could have expanded on within the individual paragraphs, but I was like 'no, I've – I've made a point, or at least the majority of the points I wanted to make within this paragraph already, it could be a lot better but this isn't a tutorial essay, I don't have time, and I think it's more productive if I just go onto the next point'.' (Nicole, Interview 2) 'I finished the essay in the sense that I got everything I wanted in, I didn't write a conclusion because of time limits.' (Phoenix, Interview 2) 'It is a shame that the essay is incomplete as an extra paragraph and a conclusion would have completed what is the start of an exemplary essay!' (Chloe, giving feedback) 'having two less sentences elsewhere would be OK – a conclusion is perhaps going to add more value than two sentences elsewhere in an already good essay!' (Gemma, giving feedback) Related to ***time and process*** under 'other codes'. References may not be coded as both.
C14. General; broad	References to the level of discussion being quite general, or including generalisations. Predominantly workshop data as general is generally used in the pejorative. Checked for consistency using a text search in NVivo. This code could be the flipside or deficit perspective on ***specific, precise***. 'The example given doesn't seem to address this fully enough to base a generalisation but generally the quotes seem precise and well

Code name	Explanation
C14. General; broad (cont'd)	chosen.' (Celine) 'The candidate has a sophisticated writing style and deals well with 'macro' ideas but lacks specific details / close analysis which could strengthen their argument' (Tessa) 'The essay becomes saturated with plot analysis at the expense of showing range – which is especially problematic, as her thesis envelops & characterises <u>all</u> Victorian writers' attitudes towards class' (Amber) Usually in the pejorative, but a few positive references in interviews to taking a broader scope: 'I think I... kind of mix well between broad, holistic, theoretical ideas' (Phoenix, Interview 2) Examples from feedback: 'The conclusion was interesting for making a broader point about modernity.' (Rachael, giving feedback) 'I felt like the question was addressed too broadly.' (Chloe, giving feedback) '…instead it is used in summary in the conclusion which leaves the essay on a more generalising note – this does not reflect the well-supported judgement in the main body of the essay.' (Poppy, giving feedback)
C15. Introduction	Any mention of the introduction – e.g. its clarity, the role it plays in the essay. Synonym: 'opening' (e.g. Megan). Like ***conclusion***, used for annotations and responses to the part of the essay – the first paragraph. 'The writer did well to explain their thesis in the introductory paragraph, making sure to explain the way they wanted to view and set up the complex terms of the question.' (Zoe) 'Introduction is very vague and doesn't answer the question. They don't outline the texts they will address so the reader has no sense of where the essay is going or any structure.' (Felicia) '<u>Successful:</u> strong introduction' (Sean)

Code name	Explanation
C15. Introduction (cont'd)	In interviews, can refer to the process of writing introductions: 'Umm yeah so like I tried to, wrote the intro, and then just like kind of checked back [laughs] and checked back at the question, I wanted to check that like I was answering that' (Megan, Interview 2) 'Umm, I often tend to write quite vague introductions, 'cause I never really know what I'm doing when I start writing an essay, so I should really plan better' (Heather, Interview 1) 'Umm. So, need some sort of intro, probably come back to that' (Megan, TAP 1) 'The introduction is very clear and gives relevant historical context and sets out the scope of the essay (texts to be covered and a sense of the argument) really well.' (Celine, giving feedback) 'The introduction is clear, precise, and sophisticated. I wonder if its length is at all to blame for the fact that you seem to have run out of time' (Amber, giving feedback)
C16. Knowledge, understanding and ideas	Comments on the writer's knowledge, and often their level of knowledge. Also refers to understanding and ideas, as these often overlap with knowledge in comments. Other synonyms: insight, familiarity, awareness, content of texts. Also inferences about how much the writer has thought about the topic. Includes genre and form – specific types of knowledge for which there are not sufficient frequencies for separate codes. References in workshop data often include praise for the writer's level of knowledge. 'I like that the writing is both elegant and intelligent: s/he's clearly thought deeply about Gaskell and this novel.' (Abbie) 'Very strong knowledge of secondary criticism shown, perhaps to the detriment of textual example.' (Claire) 'Very good essay, I enjoyed reading it & felt that the writer had a very strong handle on their material.' (Jess) 'Good knowledge of Carlisle [sic] seems out of place in inability to link it to the question.' (Heather) 'This seems to be a very strong essay, setting out an interesting response to the question/quotation with lots of specific textual evidence

Code name	Explanation
C16. Knowledge, understanding and ideas (cont'd)	and good critical/theoretical ideas' (Celine) 'They expand the implications of the analysis they conduct and achieve a high level of insight and understanding. The writer is evidently familiar with both texts, relevant literary theory and is possessed of sufficient flexibility to adapt their knowledge to meet the demands of the question.' (Holly) 'A good discussion of both content, genre, and context'. (Sean) In interviews and think-aloud protocols, includes references to memory and recall. This is often connected to revision, which is coded under ***conditions*** on the SRL coding guide. 'I also quite like questions about gender so it was, I was definitely looking at that question, but I think because it said 'women writers' I couldn't think of that many that I'd done and that I had enough knowledge about.' (Tessa, Interview 1) 'Do I remember enough of Jane Eyre to write on this question? [pause]' (Chloe, TAP 1) Also includes references to how writers used or reconsidered their knowledge: 'I think… Because I hadn't thought about it in this way before, because the question makes you rethink kind of the knowledge that you've already got umm I definitely did reconsider. Like yeah, just nuancing like what I'd been thinking about what I fundamentally was arguing or like considering but in a different way to how I would normally approach it.' (Felicia, Interview 2) 'You've got to engage with the question, so it's not just a chucking-out of everything you know – it's making it quite pointed. So, if I write 'This essay's going to argue…' I know that every time I've got to refer back to my argument and I've got to refer back to the question, rather than chucking out 'This is everything I know about Victorian women.'' (Holly, Interview 1) As with example essays data, references in feedback often focus on overall judgements of the writer's knowledge, and how they use it: 'I felt like there was a good understanding of realist texts eg Vanity Fair / Bleak House / Chadwick / Eliot but that the writer struggled to apply it specifically to the quotation.' (Holly, giving feedback)

Code name	Explanation
C16. Knowledge, understanding and ideas (cont'd)	'You demonstrate a very deep and complex knowledge of your chosen authors and texts making your argument and analysis very sophisticated.' (Felicia, giving feedback) 'The same applies to discussing Woolf's suicide – while it is good you know autobiographical info, is the conclusion the best place to introduce it?)' (Gemma, giving feedback) Throughout the data set, ***critics and theory*** and ***context*** are separate codes as they are distinct types of knowledge and reflect the data better as standalone codes. References coded at these codes are not also coded at ***knowledge*** unless they refer to another type of knowledge apart from ***context***/***critics***. An example reference coded at both ***context*** and ***knowledge***: 'I also do not think that many of the contextual points regarding Gaskell's life when necessary to developing such an argument where lots of emphasis needs to be placed on the contents of the texts themselves.' (Zoe). Related codes: ***Knowledge-question match*** (a sub-code under Standards in the SRL coding guide) Data coded at ***knowledge*** cannot also be coded as ***content*** (found under the 'Other' guide). ***Content*** is used for comments where the writer is responding primarily to the literary substance or ideas themselves in the essay, rather than considering how knowledge is constructed or what level of knowledge can be inferred from the writing. Data coded as ***knowledge*** should not be also coded as ***evidence,*** except in rare cases where the writer/reader is referring to particular evidence of their own knowledge. ***Evidence*** is more precise and refers to the presence or absence of supporting evidence, examples or detail, most often giving support for a point or argument. (This disambiguation was added following a discussion with Nadri).
C17. Nuance; complicating	References to ***nuance***, to problematising or dealing with the complexities of material or an argument. Includes the deficit perspective – e.g. superficial, stereotypes. 'This is a nice, nuanced consideration of the question that dispels the binary that an answer could easily fall into. The comparison of

Code name	Explanation
C17. Nuance; complicating (cont'd)	Woolf and Forster works nicely with this as the superficially similar, yet different when one considers their complexities.' (Joe) 'A very interesting essay, bringing together lots of ideas in a subtle and nuanced argument.' (Claire) 'Problematising approach to the novel works well' (Felicia) 'Umm I would… add another paragraph, which would nuance my argument in the way that I wanted it to go, because I definitely had another paragraph I could have used, which would have kind of made it a bit more… complex and just bring up the level, really. (Nicole, Interview 2) 'I think they want something that isn't just superficial, I think that's probably the best – something that kind of asks a question and then answers that but then asks why is that the answer? And carries on going. So I always try and - I think it is probably quite obvious from everything I am at the moment thinking of stuff, always ask why. And I think.. what they like, at least I hope that's what they like [laughs] is, that it's the amount of whys you ask and how far you can get, and asking like, getting to an answer then always being sceptical about it.' (Ben, Interview 1) 'I don't really want to write about them, I think The Waste Land is a very complicated thing, start making assumptions about it, and I don't know necessarily what it achieved.' (Heather, TAP 1) 'Can't just, basic ideas on that' (Tessa, TAP 1) 'Your introduction nuances the differences between realism and simple mimesis well. ' (Claire, giving feedback) 'The approach to the contrast between Bennet and Walcott was nuanced & lucid.' (Jess, giving feedback) Similar to ***original*** and cannot be coded as both.
C18. Original	Uses of the word 'original', or variations on this – being different, providing your own fresh interpretation. Must include 'original' or 'originality', or close synonyms, such as 'unusual'.

Code name	Explanation
C18. Original (cont'd)	'V. sophisticated, original knowledge & insights' (Jess) ''Originality' of application' (Abbie) 'I mean as it says 'discuss' it's trying to open up a discussion and so get you to think about different ideas and… put sort of lots of different ideas together so.. yeah, I think it's nice that it's quite open in the way that they are obviously interested in your personal interpretation and original ideas' (Tessa, Interview 1) 'I guess a level of originality, because you can come up with quite a lot of the same points for lots of the same authors, umm so maybe… Interesting sort of combinations or interesting ideas, but all of which are sort of justified, either why you're talking about them together, or why you're talking about that particular author in that way.' (Chloe, Interview 2) 'I think I haven't said anything that… umm [pause] it's not that it isn't original, it's just to me it seems a bit kind of… obvious in my head' (Nicole, Interview 2) '… and your points are creative: I particularly liked your idea of the "old things" as past poetic voices.' (Claire, giving feedback) 'Unique interpretation of the question, as well. Good approach.' (Amber, giving feedback) Similar to ***nuance; complication*** and cannot be coded as both.
C19. Plot	References to plot, narrative or telling the story, e.g. instead of analysing. This tends to be pejorative. 'Whilst some analysis is really interesting, particularly in the first paragraph, the majority is plot-based statements that hamper rather than develop the argument.' (Chloe) 'Sounds more like assumption - retelling story more than exploring question' (Poppy) 'Specific textual examples given in a concise manner that doesn't feel narrative, but instead relate to analytical point clearly' (Gemma) 'Unnecessary amount of plot description?' (Megan)

Code name	Explanation
C19. Plot (cont'd)	'Um, I think it deals with all the novels, um… well in terms of kind of plot and authorial intent' (Amber, Interview 1) 'In the past ones that we've looked at, I don't think, people always used to kind of stray away into talking about plot and things, but I tried not to do that.' (Nicole, Interview 2) 'In some paragraphs like paragraph 1, you give quite a lot of narrative about what happens in the text before you analyse why it is important.' (Gemma, giving feedback) 'but there is a lot of plot explication and surface analysis' (Amber, giving feedback) 'The second paragraph is very interesting but narrative based' (Tessa, giving feedback)
C20. Progresses; direction	Over the course of the essay, how the essay or argument moves along, and whether there is a direction and/or progression through points and ideas. Common keywords: progresses, direction, and development (but only when applied to the overall argument across the whole essay). 'Lack of direction, clarity, argument + structure – a more definite line of argument at outset would have changed this and allowed the essay to progress.' (Heather) 'Nice, shows diachronic progression + comparison' (Phoenix) 'Despite this, the argument is made clear in the introduction, developed well in the body of the essay and concluded by rounding up ideas.' (Celine) 'The essay seemed to open up a series of possible interpretations in the introduction, then moves to argue a point each of these. Therefore the argument does not progress, and the conclusion does not appear as a well-informed judgement.' (Poppy) 'I feel like it's got some definite direction and each paragraph shows a progression along.' (Holly, Interview 1, on her essay) 'I didn't really get to the conclusion to fully explain it, like to try and show a progression, umm so trying to take a wider view of how art and war changed.' (Felicia, Interview 2) 'I would have rather then linked it back, and then how that – how where I've strayed links back to what I said in the introduction and

Code name	Explanation
C20. Progresses; direction (cont'd)	how it's an implication, not completely something new, which I don't think is necessarily made explicit, it might be a bit confusing [laughs].' (Tessa, Interview 2) 'First paragraph lacks direction – is it light, sensuality, symbolism, subtexts?' (Heather, giving feedback) Overlapping: can be coded as ***structure*** where paragraph-by-paragraph structure is implied: 'The essay seemed to open up a series of possible interpretations in the introduction, then moves to argue a point each of these. Therefore the argument does not progress, and the conclusion does not appear as a well-informed judgement.' (Poppy) Related to ***consistency and coherence***, but can only be coded as both codes where both aspects are mentioned together. **'**Um, as I said earlier, I feel like it's quite focused. Um, it's, I feel like it's quite attached to the argument. It focuses on the demands of the question and it's linked to the quotation, so in that sense it's not just a random jumble of knowledge. It's, I feel like it's got some definite direction and each paragraph shows a progression along.' (Holly, Interview 1)
C21. Question	This is a broad code used for references in comments, responses and transcripts to the question the writer selects. For workshop data, the code is used for references to the essay's engagement with, handling of, or connection to the question – or lack of. It includes annotations on the question and its terms. *Question* also encompasses the quotation in the question as the two are often hard to distinguish. 'This essay seems very closely focused on the question and the quotation throughout' (Celine) 'Strong intro, instantly engages with the question' (Claire) Examples of annotations on the question in workshop data:

Code name	Explanation
C21. Question (cont'd)	'too narrow a range [' all significant' circled]' (Ben) 'not addressed question ['representations of war' double-underlined]' (Felicia) In interviews, the code includes includes participants' readings and/or interpretations of the question, their question choice, their use of the question during writing, e.g. for self-monitoring. 'So, I mainly focused when I was looking at the question on deciding on the different texts I would talk about. Then I was writing them down here, umm and kind of deciding what the authors are saying about idealisation. And I kind of decided that it was about innocence' (Heather, Interview 2) 'Um, I feel like I managed to… I was quite happy with the question, because sometimes the questions are quite pointed and you find it difficult to adapt your knowledge. Um, but I found it fairly easy to find a question that I thought I could answer and I quite enjoyed being able to explore my take on the question' (Holly, Interview 1) 'You know when you sort of end up writing some ideas and then you're halfway through, like don't know if this is quite relates to the question' (Tessa, Interview 1) 'So I think I was trying to, I was realising that I had to like do a bit of legwork to make the quotation [i.e. stated in the question] work for just that fact. Because I didn't want it to be like [pause] that it's general and rare and extraordinary, because that's not what the quotation was saying at all.' (Amber, Interview 2) In think-aloud protocols, ***question*** is used once the participant has decided which question they will respond to. 'So go back to question.' (Nicole, TAP 1) 'Umm so in relation to the quotation, it's stating that the power of the written word is more than simply perception…. The power of the written word lies in its grounding in nature plus past and its… generative powers [pause – writing]' (Phoenix, TAP 2) Examples from feedback:

Code name	Explanation
C21. Question (cont'd)	'It's really good that you addressed the quotation in the question title at the very beginning of your essay, making sure to consider the fascinating style in which the quotation has been written.' (Zoe, giving feedback) 'Engage with the wording of the quotation a bit more it would have been helpful to see a critic's definition of "human nature" or at least your own interpretation of it, and an unpacking of what "religious systems" and "dogmas" might refer to.' (Sean, giving feedback) 'Good focus on the terms of the question (dialogue)' (Tessa, giving feedback) Related codes for interviews and think-aloud protocols: ***Re-reading or revisiting the question*** – a sub-code of ***task perceptions*** on the SRL coding guide. These examples are also coded as ***question***. 'Outside, "'I kept the life, thrust on me, on the *outside* or, the *inner* life, with all its ample room, for will and intellect"' (Felicia, TAP 1 – italics for her emphasis) 'To discuss women's writing defining its own tradition and style… but that, I can't really think of particular traditions which I think are *unique* to women's writing necessarily, to women exclusively' (Heather, TAP 1) ***Question characteristics*** – code in the SRL guide used for interviews. Only coded as ***question*** as well where the participant is speaking about the question they selected. 'Umm… Yeah, I think that, the fact that it had something, it was a broad question but it had something specific in the quote, because we talked about engaging with the quote and so that was definitely something which I was trying to do. Yeah… Yeah, I wanted to be able to... Err, I don't know, quote it or at least be engaged with part of, different words it says, and its argument' (Nicole, Interview 2) 'You're just sort of interpreting it in that way and I think that is nice and it's attractive in a [laughing] question which just says 'discuss' because you can [laughing] say what you want to a degree.' (Tessa, Interview 1) ***Complicating the question*** – a standards sub-code used for interviews. Items also coded as ***question*** as participants discuss the

Code name	Explanation
C21. Question (cont'd)	question they selected. 'I mean, I found the quotation interesting and I thought umm [pause] that it wasn't that Clough was right or wrong but that that these two p-, that these two poets kind of interrogated and complicated what he was saying in the quotation. Which I thought might have been just like an interesting thing to write about' (Amber, Interview 2)
C22. Quotes	A type of evidence, i.e. textual evidence, rather than the quotation that forms part of the question. Refers to both use of quotations, and where quotations are lacking and could be used. Must use the term 'quote' or 'quotation'. Includes quotes from both primary and secondary material (e.g. critics). 'The direct quotes are impressive, memory-wise.' (Nicole) 'Quotes were used well and there are probably about the right number of them' (Rachael) 'Context is then embedded in textual example – HOWEVER no quotation to support' (Joe) 'Umm, I started thinking about different quotes that I could remember, umm and just sort of had notes. I didn't write them all down because obviously quotes take a long time [laughing] to write down' (Tessa, Interview 1) 'I think it went quite well, actually. Because I went back through it, and I checked [flicking through papers] because I thought I've definitely got a good range of primary quotes in there, and I've got a fair range of critics.' (Holly, Interview 2)

Code name	Explanation
C22. Quotes (cont'd)	'OK, so I could talk about Eliot's concept of like inherited literary tradition, kind of wish I knew the quote for that, but that's OK.' (Megan, TAP 1) 'If I do time, it means I limit myself to only a few quotations which are directly relevant to time, I have to manipulate some of the others umm…' (Phoenix, TAP 2) 'The quotations from secondary sources were very good' (Amber, giving feedback) 'Throughout, you cleverly interweave quotes and analysis as well as context into your argument, which is impressive.' (Claire, giving feedback) Related to ***analysis***, and can overlap – quoting texts provides opportunities for analysis. Related to ***evidence***, but should not be coded as both – to be coded as 'quote', the term needs to be used.
C23. Range; scope	References to the essay's scope and coverage in terms of the number of texts or authors written on in the essay, often accompanied with comment on how this strengthens the essay or is to its detriment. ***Range; scope*** does not necessarily refer to the number of authors - can include a range of texts written by one author. Includes examples where the writer is perceived to have covered too many or too few texts. 'However, the focus on nuanced analysis fades again: the essay becomes saturated with plot analysis at the expense of showing range – which is especially problematic, as her thesis in envelops & characterises <u>all</u> Victorian writers' attitudes towards class'(Amber) 'The writer uses a range of texts and explores their contents.' (Zoe) 'Good range of texts – but there is a need for the consistency of an overarching thesis' (Ben) 'Is being based on one text a good idea? Would maybe benefit from expanding as could discuss ideas in more depth' (Gemma) 'I suppose I could say range, 'cause I talked about quite a few different texts but then, equally, I've read more novels by Woolf and

Code name	Explanation
C23. Range; scope (cont'd)	Forster than I talked about. Also I could have written about *Dubliners* as well, so range could have been better.' (Heather, Interview 1) 'OK I'm not going to have enough time, I'll just do Woolf, Oddington and Beckett, because that's three anyway. Umm but also they didn't link as well so I'd rather concentrate on those three that do link in better.' (Felicia, Interview 2) 'Umm. I don't really know if I know, I know I don't know a range of Eliot's poetry, especially Lyrical Ballads.' (Heather, TAP 1) 'Don't really have enough authors to show a wide range of the period [pause]' (Chloe, TAP 1) 'Throughout, I thought the linking between different texts was really strong, however I wondered at points if you had almost given yourself too much to write about? The range of textual evidence is very impressive but perhaps the essay overall might be strengthened by a tighter focus on fewer texts, allowing you space to look deeper/in more detail at individual writers.' (Celine, giving feedback) '…just perhaps drop one author' (Amber, giving feedback) 'Also, whilst Eliot works incredibly well as a primary focus it is be useful to include a perspective other than his.' (Chloe, giving feedback) In interview data, includes references to the dilemma of how much to cover, e.g.: Maybe I took on one, it's always whether you take on more books and write less, less quantity or you do sort of smaller amount with more quantity. So there's always that, and I think I probably should have done one less book. But in hindsight, what can you do? [laughs]' (Chloe, Interview 1) The code subsumes *selective*, originally a separate code that emerged during interview coding, particularly in response to the case of Heather and her dilemmas about selecting material in Interview 2. *Selective* emerged both in the sense of aiming to write about a manageable amount, and only writing about the texts, points, or arguments which are most focused on the particular question. The two were merged as there is a fair amount of overlap – being selective is closely related to coverage and deciding on the appropriate range. The code was renamed from *range* to ***range; scope*** to reflect some of the meaning of *selective*. Some references were then coded as ***specific, precise***, depending on the purpose or outcome of being selective with material.

Code name	**Explanation**
	References cannot be coded as both ***range; scope*** and ***specific, precise***. The underlying idea of ***range*** is that writers need to present a range or ideal number of texts/authors. The focal idea of ***specific*** is the need for writers to be specific and precise in their analysis.
C24. Relevance	References to the relevance (or lack of) of points or discussion to the question or the argument; in interviews, some examples focus on making material relevant to the question. Near synonyms included under code: relating to the question or argument; necessary. 'Both wider context and biographical context apt + relevant' (Felicia) 'Gives really good precise examples from the text + illuminates them with critical discussion – everything feels analytical (and relevant?) rather than narrative + telling us the story of the novels.' (Gemma) 'How relevant? [s2]' (Jess) 'Make relevance clearer' (Claire) 'I think I had some critics with that which wasn't quite relevant to the question that was on the paper.' (Tessa, Interview 2) 'I think with this one, you had to struggle slightly more to make what you were saying relevant to the question, almost just by nature of the question' (Heather, Interview 2) 'True Woman on a pedestal, pedestal, as an object, is that relevant?' (Felicia, TAP 1) 'In paragraph 3 you talk about 'Eliot's inability to write a better poem' and link this to Pound & 'make it new' – maybe you could just make explicit why this is relevant to the war, and an author's investigation of the war, as it feels like you learnt this for an essay on make it new and slipped it in here.' (Gemma, giving feedback) '(and as you don't initially put his work forward as a point for discussion in your introduction, or bring him into your conclusion, I don't know if it adds much to your essay?)' (Celine, giving feedback – questioning relevance of an author being included) '…and explaining how the point was relevant to the question'. (Phoenix, giving feedback)

Code name	Explanation
C26. Signposting, topic sentences and paragraphing	Specific references to the strategies the writer uses to guide the reader through the argument or refer back to the question. Key words include signposting, paragraphing or the role of 'topic sentences', 'lead in sentences' etc. in the essay. Can also refer to key moves, such as outlining the thesis, or presenting and justifying which authors are to be discussed. Can refer to the impact of the reader on signposting, or lack of. 'They don't outline the texts they will address so the reader has no sense of where the essay is going or any structure.' (Felicia) 'The brevity of paragraphs contributes to this sense of disjunction' (Abbie) 'Intro sentence might've helped? Novels (or Forster) can't really see the big picture' (Amber) 'Perhaps at times the focus just becomes slightly muddled in exploring the ideas of time and home, which do link well to the argument, but perhaps this link just needs to be articulated more explicitly/clearly?' (Celine) 'Argument almost too apparent' (Chloe) 'Often unclear how their analysis and ideas linked to war in main body of essay – so qualifying sentence would be useful, developing the argument.' (Felicia) 'Make this the topic sentence.' (Holly) 'Could show direction of argument more clearly with sentences at end of each paragraph to tie it back to the question.' (Nicole) 'I would like to be able to be a bit clearer about where I'm going and also, like I said, towards the ends of paragraphs, just be more explicit about how that links back to the argument. I think sometimes it's kind of tailed off [laughs]' (Felicia, Interview 2) 'I – mostly it was like, OK, signal the examiner what you're doing 'cause in that kind of thing, where they need to, you need to signpost really clearly.' (Nicole, Interview 2) 'Umm [pause] yeah... umm… I think I'd also maybe do a bit more signposting midway through the paragraph, just making sure because I mean one paragraphs, you can get quite bogged down in the explanation, just making sure you pull out ever so slightly, just for like a moment before you dig back in, I think I'd do. But other than that, no.' (Phoenix, Interview 2) 'Perhaps your argument could have been made clearer and sharper by the inclusion of clear paragraph-openers to flag up where next

Code name	Explanation
C26. Signposting, topic sentences and paragraphing (cont'd)	your essay is going.' (Zoe, giving feedback) 'You could have just added a sentence to signpost to me the precise direction of your argument a tad more explicitly e.g. it is the inclusion of childhood experiences which allows an author to create a 'rounded' character rather than narrative.' (Megan, giving feedback) 'A sentence afterwards just to consolidate it and make explicit how it is relevant to the question might be useful!' (Gemma, giving feedback) 'Paragraphs could have more self-contained – sometimes the premise of one paragraph wasn't approved until the next.' (Poppy, giving feedback) No examples in the TAP data. ***Paragraphing*** and ***topic sentences*** were originally individual codes, which were merged into one and then into this one as the degree of overlap became apparent. ***Paragraphing*** here includes length of paragraphs, and paragraph structure. How material is divided between paragraphs is included under ***structure*** as it is at a more global level. Related to: ***consistency and coherence***. References can be coded as both where there is mention of how signposting does/does not contribute to ***consistency and coherence***. Signposting contributes to ***consistency and coherence***, but is not sufficient on its own. 'I reckon the essay I've just written, the umm... Coetzee and err Derek Walcott essay, umm is theoretically deeper, but as a result is a little less cogent and cohesive. Umm and I've done my best to signpost the argument but I think because the argument is a bit more complex, I - and I haven't fully comprehended how I was relating it to the question it's a little bit more woolly, I'd say.' (Phoenix, Interview 2) Readers may point to 'jumps' or 'leaps' between paragraphs or points. These are coded as ***consistency and coherence***. They are only coded as ***signposting*** when it is apparent that the reader attributes these to lack of signposting, topic sentences or poor

Code name	Explanation
	paragraphing. Also related to ***clarity;*** items are coded as ***clarity*** where it is evident that ***signposting*** has contributed to the reader's general impression of clarity. See example under ***clarity.***
C25. Specific, precise knowledge	References in interviews to the need to use specific or precise knowledge. Code developed in interview analysis. Ran systematic queries in NVivo for keywords such as specific, precise but will need to keep this code as a focus when doing final checks at the end of the coding process. 'I think more specific, focused and systematic modes of argument, supported with closely analysed textual elements, would make a more engaged essay.' (Poppy) 'I think [pause] er being original and engaging with the question and being able to draw on… quite precise or [pause] detailed knowledge.' (Megan, Interview 1) 'Because the problem with the way I learn stuff, as we've kind of discussed, is it's mind maps but they're so vast that [pause] inevitably you can't write out the whole mind map in an hour, you need to be much – what I learnt from this is that you need to be much more specific, and not try and cover all the bases, but just kind of really focus on some of them, which is what I didn't really do. I tried to get all of them, get through all of it, or at least a lot of it, and then the result was that I had to cut out quite a lot of the really interesting stuff, just to get everything done.' (Ben, Interview 1) 'So, in that sense, I was used to writing comparative essays which were not based on very formal qualities. So, when you come to uni, I think the essays that you write have got to be very specific in the sense of, what does this signify, what does this show, and how can you really link this? How can you really anchor this to the text?' (Holly, Interview 1) 'Umm, have to simplify it a lot.' (Amber, TAP 1) '…with loads of great specific references to both fiction and to criticism.' (Rachael, giving feedback)

Code name	Explanation
	Also includes 'vague' as an antonym: 'And I think for me it made it a bit more vague and looser, because I wasn't constantly anchoring back to the text.' (Phoenix, Interview 1) Similar to ***range; scope***, in which the underlying idea is that writers need to present a range or ideal number of texts/authors. The focal idea of this code is that writers need to be specific. References cannot be coded at both.
C27. Structure	References to the overall structure and/or organisation of the essay – e.g. the number of paragraphs, the order of the points, whether the authors/texts are dealt with separately, how material is distributed throughout the essay. Refers to strategies for creating a good structure – e.g. ways of comparing authors, using the question. Can refer to paragraphs where the emphasis is on how the material is structured, ordered or organised, but not signposted. 'The student clearly responds to the quotation and uses it to structure their argument, integrating it well into the introduction and conclusion' (Felicia) 'Although the use and combination of analysis and context is effective, this is undermined by the essay's confusing structure and lack of a clear argument, resulting in a messy conclusion which has to make up for it. (Chloe) 'I like that the texts are discussed through points + then similarity + difference, rather than first one and then the other.' (Celine) 'Second half is critic heavy, no critics in first half – perhaps restructuring would make it read less like they are just repeating ideas read elsewhere as they would have more space to focus on their own analysis?' (Gemma) 'The running comparison with the works of writers really effective, as he functions as context, critic and a further literary material.' (Joe) 'I still have three paragraphs. I just – it just has to have three paragraphs [laughs] because it's just the structure and I don't think I can deviate from that, it would stress me out too much [laughter] umm but apart from that, like I remember the first time we talked about it, I think I said I did quite a lot that was similar.' (Ben, Interview 2) '…and then a very loose structure of like three paragraphs of what I wanted to talk about' (Nicole, Interview 2, talking about planning) 'I just kind of ordered it in a better manner that kind of presented the thesis from the beginning, rather than kind of something

Code name	Explanation
C27. Structure (cont'd)	roundabout.' (Amber, Interview 1) 'Umm then I suppose I could do a paragraph on each text, but I think that would be a very boring format.' (Heather, TAP 1) 'I think it's best to just do like a three part, yeah, probably just intro, three paragraphs and a conclusion' (Megan, TAP 1) 'I also felt that certain points could have been reshuffled to support themselves even better.' (Abbie, giving feedback) 'It is well balanced between your three writers and well structured too – contributes to the effectiveness of the essay.' (Ben, giving feedback) Similar but distinct from ***signposting***. Coding may overlap where the reference refers to how structure is created through signposting: 'It's a lot more structured, I think... a lot clearer for the reader I hope, because I'm signposting what I'm doing, which I don't think I did before.' (Phoenix, Interview 2) 'The essay would have benefitted from having these used as topic sentences eg obscure meaning / pain in writing / silence / form to create a stronger structure' (Holly, giving feedback) Instances where comparison is used as a structuring device can also be coded as ***comparison and linking***. Can also be coded as ***consistency and coherence*** where structure is (not) used to this effect.
C28. Style	References to the style of the essay – how it is expressed, or the wording. Mainly general comments on how wording comes across. Excludes references coded under ***style – concise and precise***, ***technical – referencing***, ***technical – accuracy***. 'Awkward' (Holly)

Code name	Explanation
C28. Style (cont'd)	'Elegant last sentence' (Jess) 'Colloquial?' (Nicole) 'I really like this greatly flowing sentence with fab comparisons and contrasts segwayed [sic] in very nicely' (Megan) 'I like their style, which is sophisticated and shows an individual voice.' (Nicole) 'The assertive writing style of this essay is initially both convincing and effective…' (Chloe) 'Not the right word? [s3 – 'sincerity' circled]' (Rachael) 'In parts, the student is prone to writing in a journalistic manner: "shocking statistic", "extremely upper-middle class background" merge objective evidence with subjective opinion – irrespective of the opinions themselves, academic writing should separate the two.' (Ben) 'Umm… I think it reads quite well.' (Chloe, Interview 1) 'But then also kind of just… umm… being articulate and being able to convey your ideas well.' (Heather, Interview 1) '…but throughout, this was extremely well written, moving fluidly between texts and argument.' (Celine, giving feedback) 'I like your clear writing style at the start of the essay which makes your introduction both explicit and effective.' (Felicia, giving feedback) A few references overlap with ***voice***: 'I like their style, which is sophisticated and shows an individual voice' (Nicole)
C29. Style – concise and precise	References to 'concise', 'precise' style of writing. Includes synonyms such as 'succinct', 'pithy' and 'condensed'. 'Nice, concise opening, picks apart question and addresses its terms' (Poppy) 'Nice paragraph, condensely looks at the pros and cons of Forster's handling of India, with quotations, critics + context' (Joe)

Code name	Explanation
C29. Style – concise and precise (cont'd)	'Why change from using author's name? Feels less precise' (Gemma) 'To the point' (Sean) 'Whereas this [i.e. Prelims essay as opposed to tutorial essay] you have to be really concise and condense it because you just don't have time to keep going, like have really long sentences.' (Felicia, Interview 1) 'I'd be more succinct in my introduction, I wouldn't spend as long on it. Umm… [pause] I would, wouldn't spend as long, I'd try and, although I liked my Gerard Manley Hopkins stuff, I would try and… Make that more succinct as well, I think I waffle a bit too much about that,, I get too excited about the language.' (Phoenix, Interview 1) 'However, perhaps the introduction could have been shorter and more succinct. For example…' (Zoe, giving feedback) 'I felt some of the phrasing could perhaps be slightly sharper (eg…)' (Washington, giving feedback) '…however perhaps some of the analysis would be better in a separate paragraph, or condensed to make your final argument clearer.' (Felicia, giving feedback) 'But I love that you leapt right into the introduction without making rambling introductory parts – there's immediate engagement with the question.' (Rachael, giving feedback) The code is named in NVivo: 'Because we've also been talking about being concise and precise this term, so using active, strong verbs rather than saying, "This might signify that," it just says, "demonstrates". It's cutting out all that… sort of floweriness which you're taught at A2, long words means great, whereas now it's just, "No – be accurate, be precise and that'll do the job just as well."' (Holly, Interview 1)
C30. Technical – referencing	Where a participant points out an issue with adherence to academic referencing/citation conventions, such as underlining book titles, making sources clear, giving dates for texts. 'Lastly, the author needs to watch his/her tendency when designating book and article titles, as absence or misappropriation of such

Code name	Explanation
C30. Technical – referencing (cont'd)	designation multiple times is confusing for the reader. Once is excusable – five signals carelessness.' (Amber) 'Where are these quotes from? Useful to know' (Jess) 'Don't forget to underline titles of novels, single quotations for poems & articles- published on its own, primary texts come first, secondary text paraphrase' (Sean) 'Umm I was just thinking about little things, like putting in dates' (Nicole, Interview 2) 'I think even like on a general level where I didn't really know about underlining things, because we haven't really been taught it' (Tessa, Interview 2) 'You obviously know the order in which texts were published and give a sense of chronology but it might be good to back this up with dates if you can.' (Gemma, giving feedback) 'Make sure your style of citing novel titles is consistent' (Nicole, giving feedback)
C31. Technical accuracy	References to written accuracy, usually in terms of errors, e.g. in tenses, spelling. Includes sentence structure/syntax. Includes annotations where participants make corrections to essays as they read. 'Some of the sentences almost read like note form as clauses are in odd orders – starting with quotes etc. doesn't feel very sophisticated and means that the essay doesn't flow as we have no understanding of how the quote is relevant' (Gemma) 'This response scarcely addresses the question and is difficult to understand, with poor grammar and spelling and a fragmented writing style' (Holly) 'Quite a long sentence and I am not sure if it works. Mainly because by the time I'm thinking about Mancunian portrayal I have lost Gaskell in mind. This may just be me though.' (Megan) 'This is represented by the statistic that' (Jess – correcting wording) 'Umm I think my sentences, I'm getting better at knowing when to put full stops in, I've stopped doing sub-paragraphs and all that'

Code name	Explanation
C31. Technical accuracy (cont'd)	(Phoenix, Interview 2) 'Obviously you're writing quickly but I don't like punctuate well enough the first time around, and I don't know, umm… occasionally you read things back and you realise that you've missed a word or you've… And I think I do that quite frequently, so I don't know, I think it's a good thing for me to do.' (Megan, Interview 2) 'Throughout your essay there are a few very minor errors that you could improve by proofing – in paragraph 1 you repeat some words, in paragraph 3 you have a few errors with punctuation that make some of the orders of your phrases a bit confusing.' (Gemma, giving feedback) '…and watch out for spelling and grammar.' (Nicole, giving feedback) 'Pay attention to syntax' (Heather, giving feedback)
C32 Terminology	References to use of terminology – e.g. critical terminology, buzzwords. 'Buzz words need to be unpacked' (Jess) 'Good critical terminology' (Phoenix) 'Well identified, should develop and use this term' (Sean) 'Umm I mean I think you always kind of have the idea of the examiner in mind, so you want to obviously use quite formal language and get your terms in to show that you know something.' (Tessa, Interview 2) 'Just like making sure I use terminology that they'll like, umm [pause] just flagging like what I think mark-gaining points, so it's, cos it's I mean, a lot of the stuff the tutors have been saying, it's just making it easier for them to give you marks. And it's like a very mercenary way of looking at it [laughter] but. I'm going to try and smash and grab marks so it's just making it easy for them to see that.' (Phoenix, Interview 2) 'But I want to um, I need to choose which words I'm going to use, make sure I've got like the same keywords throughout the essay, narrative of the individual' (Ben, TAP 2)

Code name	Explanation
	'Excellent use of critical vocab: omniscience, didactic etc' (Tessa, giving feedback) 'It's good you define in what sense you are using words like 'translation' p2' (Gemma, giving feedback)
C33. Voice, positioning of the author and/or opinion	Includes references to the tone of the essay, or the way the writer positions herself/himself (e.g. Gemma: 'not assertive, or decisive'). Also, where the author asserts their opinion too much. Common terms: voice, assertive, confident, judgemental, convincing, position. Often refers to instances where the writer's position is overshadowed by use of critical material. 'There is need for a sentence or two that explains or sets out a thesis though, one that uses his/her voice.' (Ben) 'Very strong knowledge of secondary criticism shown, perhaps to the detriment of textual example. Almost used as a substitute for the writer's own argument, for example three paragraphs in a row end with a critic's viewpoint.' (Claire) 'Sometimes feels quite didactic when uses judgemental adjectives – rather than sounding authoritative they actually feel too powerful' (Gemma) 'Good that the student challenges the question (that Forster is anti-imperialist)' (Tessa) 'Again voice drowned out by critics' (Abbie) 'Convincing' (Chloe) 'Better as more their own words, less critics' (Felicia) 'Have I missed anything? [pause] oh, and you kind of have… like a strong and confident voice all the way through, like you know... where you stand or at least what you're going with it.' (Megan, Interview 2) 'Umm… I think the best ones are authoritative, so that they show that the candidate has kind of read as kind widely as possible in a week's tutorial essay umm of the topic that they're handling and that from the quotation they've understood it, they're applying it well, I mean correctly… umm and that they're kind of developing an argument from the quotation or just whatever the question might be.' (Amber, Interview 2)

Code name	Explanation
	'I felt I was uncertain of the essay's own voice at these points.' (Abbie, giving feedback) '4th para: 'it may be' – tentative, be bold.' (Jess, giving feedback) '…and original, led by personal opinions as opposed to a recycling of pre-conceived literary thought.' (Joe, giving feedback)
C34. Word count and essay length	Created in workshop data to capture references to length, and also additional annotations made on word count or length – i.e. mentioned by me in the workshops. In interview data, references include the amount the participant wrote or wanted to write, and expectations about length (e.g. in comparison to tutorial essays) 'Very short - average is 900 - this is 510' (Chloe) 'Short – Not good use of words' (Heather) 'A little brief, maybe timing was an issue' (Claire) Note that the latter example isn't coded under ***finishing*** as 'brief' appears to refer to length rather than completeness. 'So probably that just umm... trying to write as much as I can' (Amber, Interview 2) 'Yeah, 'cause I had about six minutes left over including reading, so I do tend to be able to do ten minutes planning, fifteen minutes writing an essay, and I've ended up with… a fair bit of writing [laughs].' (Holly, Interview 1) 'I think 'cause tute essays are a lot longer, so you have that opportunity to write more stuff and your tutor expects you to write more stuff, whereas I think in the examination... they don't.' (Tessa, Interview 1) Also includes instances where other standards mean a shorter essay is desirable: 'And probably tried to write a little bit less but think about it more' (Megan, Interview 2)

Code name	**Explanation**
C34. Word count and essay length (cont'd)	No examples in feedback data – all come under ***finishing*** instead. Similar to ***range; scope***. Excluding factors: References in this code focus on the length, rather than explicitly the amount of material that participants are trying to fit into the essay – number of words rather than coverage of material. No overlapping allowed.

Coding guide: other comments

These codes remain from the workshop data coding last year. They were removed from the criteria guide as they don't denote criteria as such, rather other comments that participants make. ***Time and process*** is useful for the standards chapter – suggests particular ideals about what writers should do with their hour. ***Content*** is also useful, as it shows where readers drift into making judgements on *what* it written rather than *how* it is written, showing how difficult it is to untangle ideas from their construction. Aside from ***content***, these codes are low frequency, but I am not deleting them yet as they could be useful for later coding for the Feedback chapter.

Code name	Frequency in ex essay data	Explanation
X1. Content	187	In workshop data only, references where readers are commenting on the content or substance of the argument or of the ideas discussed, rather than *how* the essay is written. Often involves the reader's interest in the material, or content-led suggestions for improving it, e.g. material or alternative positions. 'What were the social problems? [s1: 'the social problems' circled]' (Zoe) 'Really interesting + rich [s2 - whole sentence underlined; s3 – 'by mirroring these traumatic penetrations of war into peace' underlined]' (Rachael) 'Good, but romantic in what way? [s5 – 'sad, helpless, irrational, typically romantic heroine as she flees suddenly' & 'Sweetheart' underlined]' (Sean) 'I really liked your distinction between English as a language and as literature' (Claire, giving feedback) 'Could also improve arguments by discussing what a critical position seeing lit as transnational would mean' (Heather, giving feedback) Similar to ***knowledge***, but references cannot also be coded as both. ***Content*** is used for comments where the writer is

Code name	Frequency in ex essay data	Explanation
		responding primarily to the literary substance or ideas themselves in the essay; to be coded as ***knowledge,*** the reference needs to consider how knowledge is constructed or what level of knowledge can be inferred from the writing.
X2. Performance	34	Any reference in workshop data only to the reviewer's overall judgement of the writer's performance. 'A good, strong essay with a clear structure / focus and thesis.' (Phoenix) 'Really good! Really fluent and with a clear argument, critical engagement and really interesting points' (Rachael)
X3. Suggestion	3	Responses to example essays only. Early-stage coding for where students make references to what the writer should have done or could do instead. However, I will leave this to a later round as I think codes will develop that will focus on modes of suggestion, particularly once I code the feedback. 'Should think through relevance of each text before bringing them in If implications of largely irrelevant points had been drawn out / developed would have engaged well w q as its basic premise is how texts relate to their age / era etc.' (Heather) 'Lastly, the author needs to watch his/her tendency when designating book and article titles, as absence or misappropriation of such designation multiple times is confusing for the reader. Once is excusable – five signals carelessness.' (Amber)
X4. Text as an exam essay	12	Responses, annotations and additions on example essays only. References to the text being appraised as an example of the exam essay genre, e.g. highlighting the rhetorical situation / expectations. This was introduced after two rounds of coding. 'The examiners wouldn't have used a quotation if parts of it were to be "unimportant" as an afterthought in the conclusion' (Amber) 'In a paper designed to show range, I feel that two texts might also be too narrow.' (Holly) 'It is made more convincing by an underpinning of dates and statistics, but could focus more on the language of the text itself, as

Code name	Frequency in ex essay data	Explanation
		a literature essay.' (Nicole) 'More like an essay plan' (Poppy)
X5. Time and process	29	Reader-only code for references in workshop data and feedback data. References make inferences about the writer's use of the hour, the effect of timed conditions on the process, or other aspects of the writing process, e.g. planning. This code is used where participants try to figure out why the essay is a certain way – why a criterion is missing or comes across in the way it does. 'I get the impression that the student did not plan for writing which would have grouped their ideas much more effectively' (Holly) 'Is it even a conclusion or just another paragraph which they ran out of time to finish' (Felicia) 'Candidate ran out of time/panicked?' (Jess) 'It is clear time had run out to write a conclusion which is a shame.' (Chloe, giving feedback) 'I'm not sure if timing was an issue' (Claire, giving feedback) 'Bringing in a critic or two more could help to streamline your argument & show knowledge (but obviously time is a constraint!)' (Megan, giving feedback) 'It seems like you ran out of time at the end.' (Nicole, giving feedback) Includes inferences made by the reviewer about material being pre-prepared: 'Though attention to the question is given in the opening and close of the essay there is some sense of the answer being pre-prepared, though this does also mean that it is closely and concisely argued.' (Abbie, example essay data) Similar to ***finishing***. Excludes references where the emphasis is on finishing within the time.

Code name	Frequency in ex essay data	Explanation
		Equivalent writer-only codes under ***standards*** in SRL coding guide: ***time, better planning process.***
X6. Tutor interviews	3	References in workshop data – additions – which show traces of the discussion of essays where I shared what tutors had said about the essays. 'Tutors were okay with dealing with them separately.' (Ben) 'One tutor liked this?' (Gemma)
X7. Reader experience	-	Code that emerged from the feedback data to capture instances where the experience of reading is highlighted, or the reviewer emphasises their readerly perspective to the writer. '…The result being that I felt a bit like I was being taken for a ride, not knowing where I was going – but that the essay's clarity & consistency of argument throughout meant I was enjoying the ride in any case.' (Amber, giving feedback) 'These will all help convey your good points in a clearer way to the examiner/reader.' (Ben, giving feedback) '….as it feels like you learnt this for an essay on make it new and slipped it in here.' (Gemma, giving feedback) 'I interpreted the quotation very differently. I thought that Ramazani was arguing that there were certain features of post-colonial writing that was shared, across national boundaries….' (Jess, giving feedback – explaining her take on the question) '…(this is obviously a tremendously poor example – I regret writing it)' (Megan, having given an example of how to rewrite a sentence).

SRL coding guide

Overview of main codes

Phase codes

SRL 1	Task perceptions	This node refers to the phase of task perceptions. Temporally, this is from the very beginning of the task, through reading questions and deliberating over choices, up until when participants begin to plan their selected question. There is a lot of overlap between ***planning*** and ***task perceptions*** in the data – participants often return to the question.
SRL 2	Planning	The parent node refers to the phase of planning – where participants are making plans to respond to a particular question, and before they start writing the essay in full sentences.
SRL 3	Enactment	The main phase of writing. Temporally, from the first full sentence written – and when participants stop thinking aloud – until writing is finished and/or the end of the timed period.
SRL 4	Adaptation – new today	Refers to new strategies or approaches that writers report they tried in the writing session, or that they have implemented recently in their writing. Sequenced here after enactment to take into account that these adaptations refer to present writing, rather than future writing.
SRL 5	Self-evaluation	Refers to instances in the interviews where students make evaluations or judgements about the quality of their essay writing or process. Part of the large-scale adaptation phase of Winne and Hadwin model, but here are separated out from adaptation in order to distinguish between participants' judgements about their writing and/or process and their ideas or intentions for adaptation and change.
SRL 6	Adaptation – future	The final phase of Winne and Hadwin's model – large-scale adaptation. Refers to changes or adaptations that writers intend to make in future essay writing.

Cross-phase codes

SRL 7	Conditions	Any factor which may shape the way the writing unfolds. The code includes task conditions, e.g. needing to write one essay in an hour, and conditions to do with the writer her/himself: anything the writer brings to the task. This includes ***metacognitive knowledge*** of the learner as a writer (e.g. strengths, habits), previous writing experiences, confidence, amount of revision done before the task, tutor advice or feedback. The ***research situation***

		is also a condition.
SRL 8	Self-monitoring or revision	Specific instances where participants report self-monitoring their progress in writing (i.e. during enactment) or making revisions.
SRL 9	Standards	A criterion or belief that is used to judge or evaluate learning, either how it is going in the moment, or how it went (for example, in the self-evaluation part of the interview).
SRL 10	Emotion or affect	A broad bucket code where any reference to emotions, or an affective response to writing, comes up in the interview. Can also be used for emotive language.
SRL 11	Goals	Refers to evidence in the interview of participants' intentions or goals for the writing session. For example, they may have wanted to write on particular author. Unlike in the Winne and Hadwin model, goals can be set before the writing task – e.g. to use the session as part of their examination preparation.
SRL 12	TAP operations	Refers to operations in the think-aloud protocols which may take place in either/both the task perceptions or planning phases, e.g. ***self-questioning, re-reading the question.***

Each of these codes has sub-codes, and an overview is given on the next page.

Overview of sub-codes

Number	Code Name
SRL 1	**Task perceptions**
1.1	Analysing or interpreting the question
1.2	Autocoding task perceptions
1.3	Re-reading Q or revisiting
1.4	TAP task perceptions
1.4.1	Decision (making)
1.4.2	Identifying task demands
1.4.3	Narrowing down options
1.4.4	Reading through the Qs
1.4.5	→ *Demand, difficulty of Q*
1.4.6	→ *Done that*
1.4.7	→ *Interesting, like, dislike*
1.4.8	→ *Linking to particular material*
1.4.9	→ *No*
1.4.10	→ *Possible option (identifying)*
1.4.11	→ *Pre-planning, identifying direction*
1.4.12	→ *Texts, authors (linking)*
1.4.13	→ *Topic of Q (identifying)*
1.4.14	→ *Unsure*

Number	Code Name
SRL 2	**Planning**
2.1	Argument (formulating)
2.2	Clarifying
2.3	Conclusion (planning content of)
2.4	Generating ideas
2.5	Intro (planning content of)
2.6	Paragraphs (allocating material to)
2.7	Planning time
2.8	Recalling info
2.9	Sequencing
SRL 3	**Enactment**
3.1	Autocoding enactment
SRL 5	**Adaptation – new today or recently**
SRL 5	**Self-evaluation**
5.1	Autocoding self-evaluation
SRL 6	**Adaptation**
6.1	Adaptation – future
6.2	Adaptation – new today or recently
SRL 7	**Conditions**
7.1	Autocoding conditions
7.2	Collections
7.3	Examiners
7.4	Ideal
7.5	Official guidance, e.g. examiner reports

Number	Code Name
	and criteria
7.6	Revision
7.7	Tutors
7.8	Research situation
7.9	Metacognitive knowledge
SRL 8	**Self-monitoring or revision**
SRL 9	**Standards**
9.1	Better planning process
9.2	Complicating the question
9.3	Ideas
9.4	Idiosyncratic
9.5	Keywords
9.6	Knowing what you're going to say
9.7	Knowledge-question match
9.8	Question characteristics
9.9	Read over
9.10	Task perceptions as standards
9.11	Time
9.12	Understanding of the question
9.13	Use the plan
SRL 10	**Emotion or affect**
SRL 11	**Goals**
SRL 12	**TAP operations**
12.1	Judgement - e.g. of knowledge
12.2	Keyword (identifying, returning to)
12.3	Narrating
12.4	Re-reading Q or revisiting
12.5	Self-questioning
12.6	Self-talk (guiding)
12.7	Time check
	Other codes
SRL 13	Approach similar (autocode)
SRL 14	Barriers to adaptation
SRL 15	About standards

Coding guide: phases

SRL 1: Task perceptions

Code number	Code Name	Data use	Description
SRL 1	**Task perceptions**	**TAP and interviews**	**This node refers to the phase of task perceptions. Temporally, this is from the very beginning of the task to when participants begin to plan, although there is a lot of overlap between planning and task perceptions in these data.** **Auto coding on task perceptions has been carried out on the relevant questions. Many of the operations above relate to task perceptions, and will be moved under this node once the coding is done and reviewed.**
SRL 1.1	Analysing or interpreting the question	TAP and interviews	Used in both TAPs and interviews where participants intentionally try to interpret or analyse the question/quotation, e.g. by breaking it down or paraphrasing it. 'Umm after I chose the question I… wanted to check that I'd read the question right and kind of like went through it and broke it down into different like… err sections and tried to really like understand what the question was asking and like what, like almost kind of what trap not to fall into, if that makes sense [laughter]' (Megan, Interview 2)
SRL 1.2	Autocoding task perceptions	Interviews	Autocodes for easy retrieval of particular interview items. Subcodes for the following interview items***: choosing the question***; ***shifts in quotation and question***; ***unpacking the question*** and ***purpose***

Code number	Code Name	Data use	Description
SRL 1.3	Re-reading Q or revisiting	TAPs only	Code developed for think-aloud protocols to identify points where participants reread or revisits the question/quotation after their first read through. This includes re-reading just part of the question, and occurs either as they are developing initial task perceptions, or more often, in the planning phase. 'The question is: My task which I am trying to achieve is, by the power of the written word to make you hear, to make you feel- it is, before all, to make you *see.*' Write about the power of the written word.' (Phoenix, TAP 1) 'Occurs in an entranced suspension of history, as if the definitive historical events have *preceded* us' (Tessa, TAP 1)
SRL 1.4	TAP task perceptions		Codes for operations carried out during the task perceptions phase of TAP data.
SRL 1.4.1	– Decision (making)	TAP only	The point in the think aloud protocol where it appears that the participant is making a decision about which question to write on. 'I think… OK I'm going to do how does Victorian literature… treat commerce and capitalism.'(Chloe, TAP 1) 'Settling for question 7.' (Holly, TAP 1) 'Yeah, I'll do the other one' (Ben, TAP 2)
SRL 1.4.2	– Identifying task demands	TAP only	This node refers to where participants explicitly identify or remind themselves task demands in the think aloud protocol. So far, there is only one instance of this. 'Gotta write one question. [Opens paper] Right, OK.' (Megan, TAP 1)

Code number	Code Name	Data use	Description
SRL 1.4.3	– Narrowing down options	TAP only	The point in the think aloud protocol where participants narrow down the options, often comparing one question to another. This can be one brief fragment, or it can be a longer deliberation. 'Umm, that one or women?' (Felicia, TAP 1) 'So I quite like the post-colonial and…. the first one? What was the first one again? "The relationship of any modem writer to their predecessors." Can link back to the Victorian notions, sort of how did they make it new, or not, they were developing on ideas, indebted to Tennyson… [Dododo], what other ones did I like?.....' (Tessa, TAP 1) 'Come to the end of paper, got to choose one. 19. "Uses of conflict in writing", mmm, no. 20. Children's literature, no. Go back to questions highlighted, 12, 6 and 7. Do I want to write on women? Mmm, not sure…..' (Holly, TAP 1)
SRL 1.4.4	– Reading through the Qs	TAP only	Sub code under TAP operations, but also has its own subcodes to record the particular operations that participants are using as they read through the questions. This also operates as a sub-phase of ***task perceptions***, but is currently placed under ***TAP operations*** for ease of coding.
SRL 1.4.5	→ *Demand, difficulty of Q*	TAP only	Where participants comment on features of the question, in particular its particular demands or difficulty. 'Nope [laughs] That's too difficult.' (Chloe, TAP 1) 'Umm. Description, again a bit vague, would be wary about doing that.' (Phoenix, TAP 1)
SRL 1.4.6	→ *Done that*	TAP only	Where participants have already done a similar question and don't want to write another similar response. 'Done that' (Chloe, TAP 2)

Code number	Code Name	Data use	Description
SRL 1.4.7	→ *Interesting, like, dislike*	TAP only	Where participants respond to the question with their personal preferences – that they find interesting, that they like or dislike the question or similar questions. 'To be honest, I'm not fascinated, I know I'm not very interested in literary criticism and theory, it's not like my favourite thing in the world, so I probably wouldn't want to explore that relationship.' (Heather, TAP 1) 'Mm. That doesn't sound like a nice question.' (Phoenix, TAP 1)
SRL 1.4.8	→ *Linking to particular material*	TAP only	Where participants are trying to include particular material in their planning or response to a question. 'It's how to get that procreative and then to link the two together.' (Phoenix, TAP 1)
SRL 1.4.9	→ *No*	TAP only	Where participants reject or rule out a question as a possibility, sometimes simply as 'no'. 'No. Not poetry.' (Chloe, TAP 2) 'Not really anything I've done, so no'(Phoenix, TAP 2) 'No, I don't know them.' (Nicole, TAP 1)
SRL 1.4.10	→ *Possible option (identifying)*	TAP only	Where participants seem to consider a question as a possible option to answer. 'Could talk about Shaw and Dickens.' (Chloe, TAP 2) 'Umm. Discuss. Mm. That's interesting, I wonder if I could start with [?] the idea of looking back' (Megan, TAP 2)

Code number	Code Name	Data use	Description
SRL 1.4.11	→ *Pre-planning, identifying direction*	TAP only	Where participants begin to plan particular ideas that they could use in relation to a question. This node is more specific than just identifying an author, a text or a topic - e.g. it involves linking, or more elaborated ideas. 'I suppose in terms of arguing that literature is 'common ground', might be, I suppose I could write about, I don't know, the public or the purpose of literature, which I suppose kind of changed in modernism so might be a bit interesting.' (Heather, TAP 1) 'That's something I could potentially do based on religion and inertia, particularly looking at Tennyson, and then perhaps linking that to umm to Week 7, science and religion, umm.' (Phoenix, TAP 2)
SRL 1.4.12	→ *Texts, authors (linking)*	TAP only	Where participants link particular questions to authors or texts that they could write about in response. 'Truth-telling, that is uh Eliot, and realism, Gaskell, uh Dickens.' (Felicia, TAP 1) 'Yeah, you could definitely bring in Pinter, Beckett there.'(Megan, TAP 1) 'It's a really interesting one, in terms of umm George Eliot's characters, like Dorothea, Maggie… [pause] Umm [pause] where and when, yeah, I think that's a really interesting one.' (Heather, TAP 2)
SRL 1.4.13	→ *Topic of Q (identifying)*	TAP only	Where participants identify the topic of the question – either the main idea (in their view of the question), or the topics that they could write on in response. Sometimes there is overlap between this and '***texts, authors (linking***)' 'Truth-telling, that is uh Eliot, and realism, Gaskell, uh Dickens.' (Felicia, TAP 1) 'Not doing colonialism.' (Chloe, TAP 2)

Code number	Code Name	Data use	Description
SRL 1.4.14	→ *Unsure*	TAP only	Where participants appear unsure, e.g. about any aspect of the question. 'Don't know where I'd go with that' (Tessa, TAP 1) 'Not sure I'd know which bit to choose' (Chloe, TAP 1)

SRL 2: Planning

Code number	Code Name	Data use	Description
SRL 2	**Planning**	**TAPs and interviews**	**The parent node refers to the phase of planning – where participants are making plans to respond to a particular question, and before they start writing in full sentences. This is essentially a pre-writing code.** **Here, the sub- nodes are operations that participants carry out as they are planning. Most of these will probably not be relevant to the interviews.**
SRL 2.1	– Argument (formulating)	TAP only	Where participants appear to be formulating an argument or thesis: 'Right. Ah, argument. Umm... "Kept the life, thrust on me, on the outside / Of the inner life, ample room / For heart and lungs, for will and intellect / Inviolable by conventions." Umm. [pause] Confinement, restriction of women through umm concepts of womanhood.' (Felicia, TAP 1). 'Umm, probably part of the thesis is that these all seem to be love triangles and the two biggest issues caught up in the more, money and morality and that the former triumphs over the latter more often than not' (Amber, TAP 1) 'OK, I need an argument as well. OK. Umm…. [pause, writing], writers… revolt… against…

			idealisation… of women… umm… through… innocence… and.. complexity that is undermined… by such perceptions' (Heather, TAP 2)
SRL 2.2	– Clarifying	TAP only	Where participants appear to be clarifying or further defining the direction of their argument or essay. 'Where am I going? OK, so from there to there, 'it says something.' Assume a work of art is its content. Ignore the figurative bit because I'm not going to have time to write about all of that. So. OK. So the message that, so if the message is that, it can have its own meaning, not saying something of author.' (Nicole, TAP 1).
SRL 2.3	– Conclusion (planning content of)	TAP only	Where participants appear to be planning the conclusion of their essay: 'And then, <u>to conclude, uh, I think you need to say that</u> in a sense um you could also bring a bit of Barthes in here and think about these multiple interpretations. And the idea of like a reader's individual reading, umm to say that it's not to say that although the language is shared, the language is common ground, the literature has a common grounding [drinks tea] common grounding? But isn't common ground.' (Megan, TAP 1)
SRL 2.4	– Generating ideas	TAP only	Where participants appear to be generating ideas in response to the question/quotation. This goes beyond linking the question to particular texts – participants generate ideas about themes, episodes, characters etc. There is sometimes overlap with recall, but there is some link to the question/quotation. 'Could talk about Hard Times and morality, [pause] could talk about how excess of money is often linked with immorality [starts writing – 11.35], umm [pause] talk about Gradgrind in Hard Times, and Banbury, I hope that's his name, Banbury, I'm not sure, we will see, and then Silas Marner, it could be said he's [inaudible?] with his money, yet when it's stolen he's rewarded with Effie [writing, not speaking]' (Chloe, TAP 1) 'So there's conflict between the two, conflict between what, conflict between umm so the conflict in

			umm Turn of the Screw, what is the conflict between, I'm talking about ambiguity but it's also saying, does the unconscious so does the umm the unconscious of the individual, the unconscious presence [pause, writing] presence of the master Turn of the Screw, then repression, umm, [pause] so there's unconscious presence so also umm… also like kind of psychometric [???] reading of the text [pause, writing] of the text itself shows [pause] [writing] the idea of an integral individual under strain.' (Ben, TAP 2)
SRL 2.5	– Intro (planning content of)	TAP only	Where participants appear to be planning the introduction of their essay: 'So. So in your intro you just want to bring in everything you want to bring in, so you need to have a little bit of the Beckett you want to bring in, and maybe introduce a bit of Barthes if you can. Yep.' (Megan, TAP 1) 'Mmm. [writing] Introduction: construction of identity, link to quote on 'conventions'.' (Holly, TAP 1)
SRL 2.6	– Paragraphs (allocating material to)	TAP only	Where participants order their ideas or content into paragraphs: 'So the first paragraph, umm, gonna be [pause] True, True Women in the home. No. First paragraph domestic confinement, and then Princess, then Ruskin, then Gwendolen, then Mill. Umm. Second [pause] umm True Woman on a pedestal, pedestal, as an object, is that relevant? Kept my life, thrust on me. What's "ample room"? Object, and pressure of an ideal, mm, and then Aurora Leigh and education [pause], umm, and then, New Woman equals challenge, heart and lungs, what's the heart and lungs? Women and intellect? Umm. Ah Isabel, free, assonance. Umm, who else is in that section? And also Jane Eyre.' (Felicia, TAP 1) 'So if I go from there, I've got to get that from to psychomatic readings of the text itself, the jump is going to have to be to do with… psychoanalytic readings umm… umm… rev- like reviewing narratives or events, [unintelligible] is that going to have to be two? No, it's just going to be one.' (Ben, TAP 2)
SRL 2.7	– Planning time	TAP only	Where participants plan how they will use the time remaining: 'Umm it's taken 10 minutes to write a plan out, I have an hour, therefore, I want to make sure the

			intro's done in 10, maximum, start with a quotation, then want to go onto religion, that should take 15 minutes because religion and nature is premise, which is then being underlined by science, and I want to make sure that by 23 past, 33 past, I've got an hour from now, 33, I want to make sure that at 25 past [writing] I can write a conclusion. [Tears paper out, drinks tea] That's my plan.' (Phoenix, TAP 1)
SRL 2.8	– Recalling info	TAP only	Where participants appear to be predominantly recalling information – it seems well-rehearsed, already thought out. Also used where participants are recalling quotations to write down. 'And the overall arc of the question solution, no the conseq-, no the conclusion, or the thesis, is that nature was a godhead, [writing] beauty, it was a way of finding God, such as that romantic past, which means I also need to remember to get in the quotation from time Traveller, which is 'not a breath of wind was stirring', which links to stirring, links to Coleridge, Ancient Mariner, sterility, spiritual sterility, Mariner, get that quote in. [Reads back]: Conclusion: nature was a godhead, beauty, way of finding God, romantic past, this is subverted by science, so that nature came to be a double agent, we have Lyle, geology / Darwin, fossils, especially permeance of-, no not permeance, what's the word? Dust, and dirt. And that represents the coast, [writing], dust, dirt, coast, umm, it's the meeting point, [writing], and nature as symbolic life, spirit, yet becomes sterile, dead, seen at the end of Time Traveller. If nature as godhead is at the beginning, we want Gerard Manley Hopkins, umm [pause]' (Phoenix, TAP 1)
SRL 2.9	– Sequencing	TAP only	Looser than paragraphing – putting ideas into a rough sequence. 'And <u>then</u> Silas Marner, it could be said he's [inaudible?] with his money, yet when it's stolen he's rewarded with Effie [writing, not speaking] Umm, importance of money, so [<u>moving onto?</u>] independence in Jane Eyre' (Chloe, TAP 1)

SRL 3: Enactment

Code number	Code Name	Data use	Description
SRL 3	**Enactment**	**Interview**	**The main phase of writing. Temporally, from the first full sentence written – and when participants stop thinking aloud – until writing is finished and/or the end of the timed period.**
SRL 3.1	Autocoding enactment	Interview	Subcodes for: ***audience and reader, ideas that popped up, ideas written about before, shifts in quotation and question***

SRL 4: Adaptation – new today or recently

Code number	Code Name	Data use	Description
SRL 4	**Adaptation – new today or recently**	**Interview**	**Participants' self-reports of what they have tried to do differently in their essay, or ways their approach has changed recently.** 'Umm yeah, definitely was sort of linking back to the point, I was sort of trying to do that today. Umm and I think I'm still getting there with pushing things slightly further. Umm [pause] and yeah focusing on style and form, as well, I'm sort of trying to link back to as much as I can [pause] umm so yeah.' (Chloe, Interview 2).

SRL 5: Self-evaluation

Code number	Code Name	Data use	Description
SRL 5	**Self-evaluation**	**Interview**	**Refers to where students make evaluations or judgements about the quality of their essay writing or process. Part of the large-scale adaptation phase of Winne and Hadwin model, but here are separated out from adaptation in order to distinguish between participants' judgements about their writing and/or process and their ideas or intentions for adaptation and change.**
SRL 5.1	Autocoding self-evaluation	Interview	Subcodes for: ***areas for development, 10 minutes more, next time; comparing essays; how did it go; strengths.***

SRL 6: Adaptation – future

Code number	Code Name	Data use	Description
SRL 6	**Adaptation – future**	**Interview**	**The final phase of Winne and Hadwin's model – large-scale adaptation. Refers to changes or adaptations that writers intend to make in future essay writing. Often overlaps with self-evaluation.** **Often in response to the question, 'would you do anything differently next time?', but also arises spontaneously.** 'Maybe spend a bit more time planning. Like, if I had kind of outlined the thesis, I probably wouldn't have had to rewrite the introduction' (Amber, Interview 1)

Cross-phase codes

SRL 7: Conditions

Code number	Code Name	Data use	Description
SRL 7	**Conditions**	**Interviews**	**Conditions include any factor which may shape the way the writing unfolds. It can include task conditions, e.g. needing to write one essay in an hour, and conditions to do with the writer her/himself: anything the writer brings to the task: previous experiences, confidence, amount of revision done before the task. The research situation is also a condition.**
SRL 7.1	Autocoding conditions	Interviews	Coding for conditions generated by autocoding for interview questions focussed on conditions (***confidence about topic chosen, A-level and pre-university writing, beliefs about Prelims and tutorial essays, revision, workshops thinking aloud***). Code is maintained so I can trace this coding back to the autocoding process if necessary.
SRL 7.2	Collections	Interviews	A writer's previous experience of writing an essay under timed conditions for college formative exams – collections. 'When I did this essay on Friday for the collection, or that theme essay, it was a completely different question, umm but I did that question first because again it was my weakest one […] I think it's actually benefited from having the collection as well, because I was aware of where I was going to go wrong.' (Phoenix, Interview 2)
SRL 7.3	Examiners	Interviews	Writers' beliefs about what examiners expect and/or mark for in Prelims. This sometimes arises spontaneously, and sometimes in response to the audience/reader interview item or a follow-up question to do with the demands of exam writing. 'Uh yeah, I suppose you sort of write with the idea in mind that an examiner or someone… [Laughs]

Code number	Code Name	Data use	Description
			above you is going to be reading it, so you try to write in an academic fashion.' (Tessa, Interview 1)
SRL 7.4	Ideal	Interviews	Writers' beliefs about the ideal features and nature of a Prelims exam essay. These differ from ***standards*** as they are not used to regulate writing, perhaps because they are not (yet) believed to be feasible. 'Umm, my tutor always talks about the extending implications [laughs] so I think [laughs] that would have been a good idea but I think you just sort of get quite caught up in it and want to finish it off and so….' (Tessa, Interview 1)
SRL 7.5	Official guidance, e.g. examiner reports and criteria	Interviews	References to the English Faculty guidance available to students, e.g. examiners' reports, official assessment criteria and guidelines. Unlike the ***examiners*** code, materials coded here needs to contain a specific reference to the official guidance, not just general beliefs. 'Umm, so yeah, I mean I think I've like read examiners' reports before and seen so many comments about like, you're not considering the Victorian reading public, you're not considering print culture, you're not considering this, that, and I'm like, well, there's only one hour!' (Amber, Interview 2)
SRL 7.6	Revision	Interviews	References to the revision or preparation the writer did or didn't do for this task or for the forthcoming exams. 'I had a real struggle last week, because I've been trying to plan a condition of England essay for about five days, and it just wasn't going into my head. So I sat down yesterday and just forced myself to write a timed essay, and that came out OK, but I think now I've written that, I knew what I did wrong' (Holly, Interview 2)
SRL 7.7	Tutors	Interviews	References to college tutors' advice or beliefs about writing.

Code number	Code Name	Data use	Description
			'I used to sort of write very long paragraphs at the beginning of term, and my tutor had said to me, like, you know, "These are slightly different points. You can just write two paragraphs."' (Heather, Interview 1)
SRL 7.8	Research situation	TAPs and interviews	Where the research situation becomes particularly apparent. Examples in TAPs: where a participant asks the researcher for a time check; a reminder to keep talking. Interviews: any indication that it's not the 'real thing' – the exam – is particularly apparent. There is automated coding for the research situation – this is just the interview item at the end of the schedule on the perceived impact of the think aloud on participants' writing processes. '[I remind to keep talking] [laughs]' (Nicole, TAP 1) 'Probably I just thought because I was aware I was being recorded, I probably thought more about the question than I might have done otherwise. I think I might have jumped into things a bit too soon, otherwise. Just as you're, because you, even though you don't have to explain yourself because you sort of felt like you had to sort of somehow justify what was coming out of your mouth' (Chloe, Interview 1) 'Maybe in a Prelims I wouldn't have done this, it was a bit of a high risk not talking about moralists so openly, but like I'm not sure there's a specific set of writers that are the Victorian moralists.' (Ben, Interview 2)
SRL 7.9	Metacognitive knowledge	TAPs and interviews	Participants' declarative knowledge of themselves as learners and writers, e.g. strengths and weaknesses; habits and foibles. (Harris et al., 2009). 'Yeah, 'cause I know that I struggle with conclusions so, when I'm writing an introduction, I constantly have to think, "Right, how am I going to develop this? How am I going to round this off but show that I've made progress?"' (Holly, Interview 1)

Code number	Code Name	Data use	Description
			'Could do gender but quite good on that.' (Chloe, TAP 2) 'Umm [pause] But saying that, I know me and I also want to try and fit in as much as I can, so I might be stupid and try and put in another point. But I hope by now I realise that that's not a good idea.' (Phoenix, Interview 2)

SRL 8: Self-monitoring or revision

Code number	Code Name	Data use	Description
SRL 8	**Self-monitoring or revision**	**Interview**	**Specific instances where participants explicitly report self-monitoring their progress in writing (i.e. during enactment) or making revisions.** 'As I went along, as I finished a page, I would kind of look over that and see like how it was progressing' (Amber, Interview 2). 'I remember I kept going back and making sure that… like… I was keeping the right keywords [pause] throughout' (Ben, Interview 2)

SRL 9: Standards

Code number	Code Name	Data use	Description
SRL 9	**Standards**	**TAP and interview**	**A criterion or belief that is used to judge or evaluate learning, either how it is going in the moment, or how it went (for example, in the self-evaluation part of the interview).** Examples from TAPs: 'Don't really have enough authors to show a wide range of the period [pause]' (Chloe, TAP 1) Inferred standard: show a wide range [of literature] of the period. 'Umm True Woman on a pedestal, pedestal, as an object, is that relevant?'(Felicia, TAP 1) Inferred standard: relevant to the question/argument 'OK, so I could talk about Eliot's concept of like inherited literary tradition, kind of wish I knew the quote for that, but that's OK.' (Megan, TAP 1) Inferred standard: use quotations, but this is adjusted for the writing situation (not an exam, no revision). 'Need to remember critics. Try and get lit theory in if I can.' (Phoenix TAP 1) Inferred standards: use critics, use literary theory if possible. Using critics seems pretty essential; literary theory seems desirable. Examples from interviews: 'I think I spent too long on my introduction, which I didn't mean to do. Umm I like my introductions to like, be like my conclusion, to encapsulate my argument, so that if I don't get to a conclusion

Code number	Code Name	Data use	Description
			because of time… They know what my argument is and what my contention is. But I spent too long focusing on that, which meant that I was rushed towards the end and didn't quite get to the grit that I wanted to in the end. So, yeah. And I didn't get enough critical analysis in, either… But yeah. I got the premise of my argument and everything in, so that's OK.' (Phoenix, Interview 1) Inferred standard: state premise of argument in introduction. This is not only an ideal ('I like my introductions to…') but a standard as he states he 'spent too long focusing on that'. He also uses this standard in his self-evaluation: although he didn't 'get to the grit that I wanted to' he did succeed in that he 'got the premise of my argument and everything in, so that's OK' 'I did add in like an extra little sentence, <u>because I thought I hadn't made myself clear enough at one point</u>' (Megan, Interview 2) Inferred standard: clarity 'So I think I was trying to, I was realising that I had to like do a bit of legwork to make the quotation work for just that fact. Because I didn't want it to be like [pause] that it's general and rare and extraordinary, because that's not what the quotation was saying at all.' (Amber, Interview 2) Inferred standard: here, Amber's task perceptions operate as a standard: she does 'legwork' to remain true to her interpretation of the quotation – 'I didn't want it to be like [pause] that it's general and rare and extraordinary, because that's not what the quotation was saying at all'.
SRL 9.1	Better planning process	Interviews	Instances in interviews where the standard seems to be 'better' planning. Sometimes 'better' is fuzzy. 'Um, maybe spend a bit more time planning. Like, if I had kind of outlined the thesis, I probably

Code number	Code Name	Data use	Description
			wouldn't have had to rewrite the introduction, so… yeah.' (Amber, Interview 1) Inferred better planning process: more time pre-writing. 'I think I could have had more things I could have probably put in if I'd possibly planned a lot better, but…' (Chloe, Interview 1) Inferred better planning process: unclear. Could be: plan to generate ideas and see how they fit? 'So, yeah, I think, in exams, I really need to decide what I'm saying to be able to then kind of push my ideas, but I sort of think, "Oh, I don't really know what I'm gonna say. I may as well begin." And then, by the time you're half-way through an essay you're like, "Oh, OK, cool. I'll talk about that then." [Laughs].' (Heather, Interview 1). Inferred better planning process: make a decision before starting to write, rather than letting ideas emerge through writing. 'Mm yeah, I think it went OK... Umm [pause] I think [pause] I hadn't like, I think I picked an OK question but… umm… I think if I'd just spent like an extra minute planning, I think I could have got... a better argument, because I err was using Coetzee for my argument, I could have developed a better argument about that…' (Megan, Interview 2) Inferred better planning process: spend more time developing argument
SRL 9.2	Complicating the question	Interviews	Instances where students describe taking the ideas in the question and complicating or challenging them. 'Mmhmm. I mean, I found the quotation interesting and I thought umm [pause] that it wasn't that

Code number	Code Name	Data use	Description
			Clough was right or wrong but that that these two p-, that these two poets kind of interrogated and complicated what he was saying in the quotation. Which I thought might have been just like an interesting thing to write about.' (Amber, Interview 2) 'I didn't really write about, I don't know, I didn't write about in the way, because I could tell it was about imagism and like... umm aesthetic modes and-, but I kind of umm turned it towards war [laughs] and I don't know if that works or not, but I gave it a go [laughs]' (Felicia, Interview 2)
SRL 9.3	Ideas	Interviews	Code used where ideas in the question/quotation are a key reason for choosing it. 'So I thought that's quite an interesting idea that links to, umm, lots of things like *Tess of the d'Urbervilles* where Angel Clare obviously sees her as like this ideal beauty and doesn't realise that she's like just as, like, flawed as anyone else.' (Heather, Interview 2)
SRL 9.4	Idiosyncratic	Interviews and TAPs	Standards that seem particular to the individual, and don't fit with other codes. 'Knowing how much I can say in the paragraph, I think. It's like you can just say about two points, like *I* can, I don't know what everybody else does but I find that I can only do two points like in the depth that I need to do them in. ' (Ben, Interview 2) 'I think it's a sonnet, but I haven't really learned any formal features yet, that's probably something I should do. [pause - writing]' (Heather, TAP 2 – standard = learn formal features)
SRL 9.5	Keywords	Interviews	Instances where keywords are used as a standard, e.g. for choosing a question or for ensuring consistency. 'Yeah. So the first thing I do, as I think you will hear from the recordings, I'm looking for key words. So, I know that I am not so strong on poetry as I am - I'm quite strong on novels. Anything with

Code number	Code Name	Data use	Description
			poetry or anything with theatre that I haven't studied goes straight out, and I narrowed it down to three which had my key words, which were like science, realism and… I can't remember, I inter-, I interpreted it to be like a fem-, a feminine reading.' (Holly, Interview 1) 'Oh, I remember I kept going back and making sure that… like… I was keeping the right keywords [pause] throughout, so I remember at the beginning I was making sure like in the plan thing I think I'd think I'd at the very end said or at least I did hear, I wrote down some keywords that like, the buzzwords that I wanted to use which [pause] yeah, and then I made sure [office phone ringing] that the introduction was linked to that.' (Ben, Interview 2)
SRL 9.6	Knowing what you're going to say	Interviews	Where participants feel they need to have a clear idea of what they are going to say or argue in their essay. 'Planning is just me getting my ideas down and sort of understanding where do I want to go, what I think about this and how am I can write it. But not necessarily exactly what I'm going to write' (Tessa, Interview 2) 'Umm well I've kind of since like the classes and everything we had and stuff I've had at college, I've definitely started just planning more, sort of 10, 15 minutes just so I have it, because you need to have an absolute idea of what you're going to say, rather than just going in. I can't just go into it and start writing.' (Ben, Interview 2)
SRL 9.7	Knowledge-question match	Interviews and TAPs	A key standard for choosing a question – the question links to or allows participants to write about knowledge they have developed before. 'I chose question number 8 so about 'how modern writers challenged assumptions about the 'content' of art'. And I think it's erm … [pause] I don't know, I think when, when you read like a quotation and something kind of jumps out at you – about what you-you've read before – and it kind of gets er… [pause] it-it gets your thinking going and it's like, 'okay, how can I apply the information I've already learnt to a question?'' (Nicole, Interview 1)

Code number	Code Name	Data use	Description
			'Umm given I don't want to do that, still want something to do with nature given that umm focussing on Hopkins, let's look for one that looks, look for one that... that discusses nature or religion.' (Ben, TAP 2)
SRL 9.8	Question characteristics	Interviews	Where the particular characteristics of a question (open, closed, emphasis etc) come into play as standards for question choice. 'It's quite an open question, like discuss Victorian attitudes to nature. That's just, that's – there's no real kind of focusing element in that. It's so broad, it just opens it to be quite a useful question to explore.' (Ben, Interview 1) 'So for the industry questions I come up with individualism, so I thought OK, I could twist that into industry, but that's a bit of a tricky one so I went, I might not do that one. I saw one on realism and thought 'ooh, do I want to do realism?' But then these questions were quite… I don't know, they were an odd twist, 'where do you find Victorian writing characterised by *such* realism?' It's - I wasn't sure.' (Holly, Interview 2)
SRL 9.9	Read over	Interviews	Where participants talk about wanting to read over their work before time is up, usually to proofread or to check for coherence and consistency. If Tessa had 10 minutes more: 'So I think definitely just reading over it, and proof reading but particularly looking at the conclusion and the introduction and seeing how they relate.' (Tessa, Interview 2)
SRL 9.10	Task perceptions as standards	Interviews and TAPs	Where task perceptions, e.g. interpretations of the question, are set as standards or used as standards during writing. 'I did to try and get the first paragraph to sound a bit more like in keeping with the ideas of, or with

Code number	Code Name	Data use	Description
			the with the terms of the question suggests' (Ben, Interview 2) 'Questioned the common ground as I was writing and also um [Pause] yeah. [pause] Questioned the, yeah the concept of private ground as well. So kind of like questioned the terms more as I was writing.' (Megan, Interview 1) ''Discuss.' Discuss, I can take that where I want, discuss.' (Tessa, TAP 1)
SRL 9.11	Time	Interviews and TAPs	Where time is used as a standard to manage writing. 'Erm just went straight into yeah, writing the introduction, I didn't really go back and change anything as I went along, umm and like I kind of scribbled down a few things, changing my plan as I was writing, I was like OK, not going to have time to talk about umm Elizabeth Gaskell, for instance, so I'm just going to tailor it towards one author, talked about three authors, but umm [pause] yeah, just tried to, tried to keep a good pace, so I – yeah, time is definitely an issue for me, so I was trying to… not fall in the trap of you know, spending too long right at the beginning. I think I might have spent a little bit too long on the introduction [pause] but yeah, I think it's good to just have – just try and write, because even if takes a little bit, I think it's better to have a clunky introduction and then you get into it, rather than just not finishing at all.' (Nicole, Interview 2) 'And then making sure that I [whispers] have time for a conclusion.' (Amber, Interview 2) 'Ignore the figurative bit because I'm not going to have time to write about all of that.' (Nicole, TAP 1)
SRL 9.12	Understanding of the question	Interviews	Where understanding the question is a standard for choosing an item to write on, or checking understanding is a standard in a subsequent phase.

Code number	Code Name	Data use	Description
			'And then I kind of wrote the title out again just to check that I wasn't on the [Laughter] wrong track. [Laughs]' (Megan, Interview 1). Inferred standard: have good understanding of question before starting to write 'Yeah, and… [laughs] first off I was kind of a … trial of process and error and like, 'okay, I don't know what this quote is really getting at'' (Nicole, Interview 1) Inferred standard: understanding the question/quotation as a key standard for choosing 'I also quite like questions about gender so it was, I was definitely looking at that question, but I think because it said 'women writers' I couldn't think of that many that I'd done and that I had enough knowledge about. So… and also I – I don't know if I thought… I was maybe a bit confused by the wording of the question and stuff and so I just sort of discounted it.' (Tessa, Interview 1) Inferred standard: be sure / clear about question wording)
SRL 9.13	Use the plan	Interviews	Where using the written plan is a standard for enactment – the standard tends to be not to deviate from the plan. 'It was a bit annoying that I couldn't – because I wanted to make sure that I was sticking to the plan, because that was also quite important in an exam situation. So I kind of had a few sentences where I just go off and talk about that and have to make sure I kind of, go back and make sure [laughing] I stick to the plan for the rest of the essay.' (Ben, Interview 1)

SRL 10: Emotion or affect

Provisional number	Code Name	Data use	Description
SRL 10	**Emotion or affect**	**Interviews**	**A broad bucket code where any reference to emotions, or an affective response to writing, comes up in the interview. Can also be used for emotive language.** **Paralanguage, e.g. laughs and sighs, are not currently coded as emotion/affect.** 'I've been really enjoying like the free writing thing' (Megan, Interview 2) 'So it's, I know myself and a few other people get quite worried sometimes about trying to get that into an exam.' (Phoenix, Interview 1)

SRL 11: Goals

Provisional number	Code Name	Data use	Description
SRL 11	**Goals**	**Interviews**	**Refers to evidence in the interview of participants' intentions or goals for the writing session. For example, they may have wanted to write on a particular author, or use the session as part of their examination preparation.** 'Um, I think my problem in collections is that I tried to write several of them on multiple authors, which resulted in kind of spotty treatment of each of the primary texts. Um, so I guess my main aim was to try out writing on one author and kind of digging into… each novel. Um… so, just trying that out.' (Amber, Interview 1)

			'I think it was just because in the last couple of days I've been revising certain things, and is I know this at the moment, and I haven't actually revised this bit yet, it's just sort of vague knowledge that I have, but I thought again it was kind of like I want to see where I'm gonna go with that, what do I know so far, and so I thought that was useful.' (Tessa, Interview 2)

SRL 12: TAP operations

Provisional number	Code Name	Data use	Description
SRL 12	**TAP operations**	**TAP only**	**This node refers to operations in the think-aloud protocols which may take place in either/both the task perceptions or planning phases** **At present, these codes have only been applied to the TAPs – it seems they will not be relevant to the interviews.**
SRL 12.1	– Judgement - e.g. of knowledge	TAP only	Where participants make judgements, e.g. about the amount of knowledge they have, and whether it is 'enough' for the question. 'Umm, that is quite a nice question. Err, I could definitely do some stuff on that, I know quite a bit about Eliot, so I could definitely have a stab at that.' (Megan, TAP 1)
SRL 12.2	– Keyword (identifying, returning to)	TAP only	Where participants identify keywords in the question, and in some cases continue to refer to these in the planning phase. 'I think my best two options are literature and science [pause] or [turns over the page] representation of money and morality in writing of the period. Money and morality, I could look at Jane Eyre, and Silas Marner, [pause]' (Chloe, TAP 1)
SRL 12.3	– Narrating	TAP only	Where participants seem to be telling what they are doing, rather than simply verbalising it –

			narrating it almost for the researcher's benefit, or perhaps their own. In traditional think aloud research, this would be a sign that the processes interrupting cognition, and that the quality of the protocol is compromised. However, the insights from these are quite interesting, and I'm not interested in 'pure' native processes but rather metacognitive awareness. 'So opening the paper.' (Holly, TAP 1) 'Gosh, there's a lot of questions.' (Heather, TAP 1) 'So we have umm Hopkins, Hopkins, I'm like drawing a mindmap…' (Ben, TAP 1) 'Umm just looking at buzzwords.' (Phoenix, TAP 2)
SRL 12.5	– Self-questioning	TAP only	Any instances of participants asking themselves a question, e.g. to question knowledge. 'And which critics do we have for him? [pause] I'll think of some.' (Phoenix, TAP 1) 'What other novels can I do? If I change that to class, I could add Wilde in.' (Chloe, TAP 2)
SRL 12.6	– Self-talk (guiding)	TAP only	Where participants talk to themselves, e.g. using the direct address of 'you', to guide themselves through the process. [Thinking aloud about Eliot..] 'And is being annoying, don't put that in your essay.' (Megan, TAP 1) 'I should probably write down the character name [inaudible] Merton Densher… (Amber, TAP 1)' 'How am I going to start it? Start off with nature and talk about, so it's attitudes to nature…' (Ben, TAP 1)

SRL 12.7	– Time check	TAP only	Instances where participants check the time to monitor their progress. 'What time is it? Half past. Umm. [pause]' (Felicia, TAP 1)

Other codes

SRL 13	Adaptation – approach similar (autocode)	Interview	Coding for adaptation captured by the interview item 'did your approach today differ to how you'd usually do a timed essay?' Helpful for looking both at adaptation and the impact of the research situation.
SRL 14	Barriers to adaptation	Interview	Instances in interviews where participants discuss why it might be difficult to make changes to the way they write or prepare for exams, or what makes them reluctant to follow advice given in feedback or elsewhere. 'Yeah, it's probably quite – it takes quite a long time to learn all these mind maps obviously. And it's getting worse and worse as it kind of goes up a level from like GCSE upwards. […] But it's kind of, it's the way I've done it, for obviously for the last four years now. And it's too, I couldn't – it's too dangerous to like change it now. I just kind of have to stick with it and, but it works well, it's obviously worked quite well, well enough at least.' (Ben, Interview 1) 'Umm… Quoting critical sources. [pause] I think, I mean I might have done it a couple of times, but there are some essays that we looked at where kind of what critics thought was *really* delved into. Umm… And it was just something that I've always not been able to incorporate, given the time constraints. And so, I guess just given [pause] a few minutes, I would take the opportunity to do that.' (Amber, Interview 2) 'Think it confirmed ideas that I already had. So like I kind of have, it's kind of sort of built on stuff already, it has made a difference, it's just that there wasn't anything specific that I took and then

			acted on, if that makes sense.' (Felicia, Interview 2)
SRL 15	About standards and criteria	Interview	Instances in interviews where participants talk about standards and criteria, rather than evidence that implies use of standards. 'In the exam you obviously have to reduce yourself to some basic critics that you can remember, otherwise you're never going to remember everything.' (Tessa, Interview 1) 'I'm not sure – is it, is it better to be, to stick to like a criteria or to try and… I think you're supposed to be able to do both, but we never really talked much about having really interesting, different ideas, but maybe if you can just have your own, not be actually seeking that, but just… be original and write what's interesting, I think that would come across, if you find it interesting in your essay, while sticking to all the kind of criteria we have of writing style, and signalling what you're doing, timekeeping and all the rest of it' (Nicole, Interview 2)

Appendix K: Layers of task perceptions

Explicit **Anything that can be identified from the task on the page**	**Implicit** **Reading between the lines about what the question / quotation is about or what is required in this examination**	**Socio-contextual** **What is valued specifically in the Oxford English Faculty context** **Literary practices and values**
Identifying the topic of the question ('it's asking to discuss the representation of money and morality in the writing of the Victorian period.' – Amber, Interview 1. Is this even codeable?) **Or the key 'buzzwords' / issues** (Phoenix, Interview 2)	**The need to use the key words of the question in the answer** ('**I sort of made sure I highlighted like 'representation' and just key words and things like that.** So **I was just constantly aware of what I needed to be saying and rather than just sort of going off one [laughs**].' Chloe, Interview 1)	**Originality** ('**try and sort of reconcile what you… maybe think about that, like personally, maybe try and get some original [Pause] viewpoint into it, if that makes sense.' Megan, Interview 1)** **'I think they probably quite value it, if you're a bit of original with your argument' Chloe, Interview 1**
Identifying the key word in terms of action the question requires ('so I think the keyword in this one is 'how successfully does it describe?' It's asking you to make a judgement and err make a sustained argument with evidence for that' – Holly, Interview 2)	**Highlighting the skills / concepts required to carry out that action –** 'make a judgement' ('so I think the keyword in this one is 'how successfully does it describe?' It's asking you to make a judgement and err make a sustained argument with evidence for that' – Holly, Interview 2)	**Related to this – writing about something interesting, or seeing something interesting in the quotation / question?** (See Amber, Interview 2)
Identifying instructions (e.g. 'gotta write on one question' – Megan, think-aloud 1)	**Genre features in terms of structure – e.g. the need for an introduction and a conclusion** (''need some sort of intro, probably come back to that, I think it's best to just do like a three part' – Megan, TAP 1)	**Literary thinking – e.g. through theory** ('**kind of almost what it means to be writing in the postcolonial umm and how [pause] like you approach that**,' Megan, Interview 2)

Explicit	Implicit	Socio-contextual
Paraphrasing the question and its demands (‘I think it’s asking me [Pause] to discuss… whether **I** think literature has a… or **from the literature that I’ve read**… **whether it has um a place in sort of a private ground or whether it is a… common shared thing. And how would you deal with those two aspects?**’ Megan, Interview 1) ‘I guess, if morality and money are developing in a certain manner in the Victorian period, how is that reflected in the fiction?’ (Amber, Interview 1)	**Needing an argument** (‘**I’ve been asked to form my own argument, which is ‘discuss’**’ – Holly, Interview 1)	**Interpreting the question in a way that aligns with what faculty/ examiners value. Or is this implicit as she cites the examiners’ reports?** ‘I kind of like the interpretation of the question – like, connecting it to the status of women – just because I know it’s something… **a historical concern in the period and I know, like, reading examiners’ reports on this paper, one of their primary concerns is that people aren’t connecting cultural and historical concerns enough’ (Amber, Interview 1)**
	Where questions don’t explicitly instruct students to comment on the quotation, recognising the need to respond to the quotation as well as the question **(‘**‘I’m never sure whether to refer to the quote in the question **and in this case I decided not to, which probably was not a good idea’ Chloe, Interview 1)**	**Literary values – e.g. Intertextuality / dialogue between texts; nuance and complication** (‘I found the quotation interesting and I thought umm [pause] that it wasn't that Clough was right or wrong but that that these two p-, that **these two poets kind of interrogated and complicated what he was saying in the quotation. Which I thought might have been just like an interesting thing to write about.’ Amber, Interview 2)**
	Awareness of the paper as a *period* paper – ‘**it’s showing your knowledge of the period and why, rather than just a specific topic.** ‘ (Chloe, Interview 1)	
Where the question asks students to comment on the quotation, identifying this instruction and operationalising it (‘I think **comment on Clough's quotation**, you have to kind of **assess the truth value of it**, and then if they're asking you to talk about more than one poet, **surely they're looking for a comparison or a contrast in relation to the quotation**’ Amber, Interview 2).	**Unstated skills / purposes:** ‘my guess just to highlight and show your knowledge of the texts…’(Chloe, Interview 1)	**What is valued in literary writing** (‘my guess just to highlight and show your knowledge of the texts, and to be able to pick out probably issues that maybe seem minor in comparison to more fundamental aspects of the texts.’ Chloe, Interview 1)

Explicit	Implicit	Socio-contextual
Paraphrasing the question - simplifying it: 'I think **just** to explore like the **representations of capitalism in Victorian literature, so if it's treated as something positive or something negative' (Chloe, Interview 2)**	**Linking the question to key literary concepts that could be addressed:** 'I think just to explore like the representations of capitalism in Victorian literature, so if it's treated as something positive or something negative, or… yeah, **how that, how it relates to the situation in the novel and to the audience, or the society at large**, I think.' (Chloe, Interview 2)	
	Thinking – both demand, and the literary practice of going deeper: 'I think it's asking you to **challenge surface perceptions,** because I did start off with, OK, you know, women are accepting of this identity, but you've got, it's not something you can take at surface value. It's only four lines, but there's so much within those four lines. **So, it is asking you to read deeply and to challenge, I think'** (Holly, Interview 1)	

Conditions that feed into task perceptions:

- Preference for type of question (e.g. Chloe's aversion to and Holly's preference for more open questions)
- Beliefs about the purpose of the question, and the function of the quotation
- 'Confidence' about the topic, and linked to this, how much revision and what type. Which topic it is out of their three – e.g. the weaker topic (Phoenix)
- Writing to discover (e.g. Megan, Interview 1) vs writing to use knowledge (Ben, Interview 1; many Interview 2s).
- Whether they come with a topic in mind, or whether they'll write on anything interesting
- Strategies for essay writing that they've brought from school (e.g. Holly – question as the third text; Megan abandoning the three paragraph structure and changing her planning)
- Emotions – finding exams very stressful (Chloe); enjoying essay writing (Holly)

Appendix L: Feedback coding guides

1. Type of feedback comment

- The guiding question for this coding is: What sort of feedback is the participant giving in this end-comment / annotation?

- Examples (in blue) are given for each code. Unlike the criteria coding guide, coding MUST be exclusive. That is, each segment of feedback can only be coded under one code. For example, a segment CANNOT be coded as ***problem*** AND ***other-dialogic***. The only exception is for using ***localisation***. For example, data can be coded at ***problem*** AND ***localisation*** if the feedback-giver both points out a problem and specifies where this is.

- Comments and annotations are also coded using the criteria coding guide. The feedback coding guide is applied first, to segment the data. Once each feedback comment or annotation is segmented into ***problem*** or *solution etc,* the content of the comment is then coded using the criteria coding guide.

- Recipients' annotations on feedback, e.g. expressing dis/agreement, are also coded as for the original comment to which they respond (i.e. for feedback segments and criteria, so if a participant agrees with a solution offered on argument, their annotation will be coded as solution and argument, too).

Code name	**Explanation and examples**
FB1. Praise	A comment that describes a positive feature of the essay (this definition of code as for Patchan et al., 2016). 'As I mentioned above I thought that the textual detail was great and well engaged with' (Abbie) 'The introduction is very clear and gives relevant historical context and sets out the scope of the essay (texts to be covered and a sense of the argument) really well.' (Celine) 'I really like the fact that you made a point to address the quotation at the very beginning of your introduction.' (Zoe)
FB2. Problem	Description of something that the reviewer **explicitly** considers to be wrong, **or detracting from the quality of the essay.** Code definition adapted from Patchan et al., 2016 – feedback givers rarely consider something to be *wrong*. Problems are often expressed directly: 'Sometimes the links between texts and paragraphs were tenuous.' (Joe) 'It is a shame that the essay is incomplete' (Chloe) '…however failure to structure your argument from the outset renders these features less convincing' (Heather) 'In some paragraphs like paragraph 1, you give quite a lot of narrative about what happens in the text before you analyse why it is important' (Gemma)

Code name	Explanation and examples
	However, they are sometimes hedged or phrased as the reader's interpretation: 'I would be wary of opening paragraphs with critical opinions' (Abbie)
FB3. Solution	An explicit suggestion for how to fix a problem or improve the quality of the essay. Usually explicit, but can be expressed in terms of 'could have' – past conditionals – when an action can be easily deduced. Code adapted from Patchan et al., 2016. Participants in this study were not suggesting immediate revisions, so past conditionals are more common. 'Engage with the wording of the quotation a bit more'. (Sean) '…but you need to be a bit clearer about why you are using it' (Tessa) 'Perhaps consider simplifying some of your sentences (for example……' (Zoe) 'but perhaps the essay overall might be strengthened by a tighter focus on fewer texts, allowing you space to look deeper/in more detail at individual writers.' (Celine) Solution: focus on fewer texts so you have space to look at writers in more depth/detail Examples with past conditionals: 'Could have used more historical context and maybe one more critical viewpoint in order to strengthen the argument.' (Chloe) Solution: use more context and one more critic '…which you could have elaborated on to make this point even stronger' (Claire) Solution: elaborate on this point 'Perhaps in your conclusion a direct comparison between two of your chosen authors could have been explored in more depth, however this would only be a suggestion, rather than a criticism of the conclusion.' (Ben)
FB4. Justification for solution	A comment that explains the benefits of a solution for the quality of the writing. Emerged from the data. Only used where a justification for a solution is given in a separate clause to the solution itself. 'Plan your time so you can write a conclusion! <u>Everything you say is good, so having two less sentences elsewhere would be OK – a conclusion is perhaps going to add more value than two sentences elsewhere in an already good essay!'</u> (Gemma on Nicole's essay)
FB5. Problem-solution	A comment that implies an aspect of the writing is problematic, or detracts from the quality, and gives an implicit solution or alternative approach. Often uses past conditional – e.g. what the writer could have done. Emerged from

Code name	Explanation and examples
	the data, and used for few examples where the problem and solution cannot be separated. 'However, bringing such a variety of writers together limits the space you have to elaborate on their positions which would help within even closer engagement with the question.' (Abbie) Problem: too many writers, limited space to elaborate Implied solution: use fewer to elaborate on positions and engage with question more closely 'Perhaps if the introduction had been more simple – the views of critics could be removed and the rest of the essay would flow a lot better, and would have been easier to write.' (Chloe) Implied problem: Introduction too complex and means rest of essay doesn't flow well Implied solution: simplify introduction, remove critics' views
FB6. Localisation	A comment that describes the location of the problem or, less often, the positive feature. Patchan et al. only use this for criticism. 'Your quote about art and modernity in paragraph 3 feels a bit unexplained' (Gemma) 'For example, the consideration of fragmented narration in the second paragraph seemed to introduce so some elaboration could have been nice' (Abbie) 'In your fourth paragraph it would have strengthened your point further about free indirect discourse to include more examples, but this is a minor point!' (Claire)
FB7. Value judgement	A comment that gives the reviewer's judgement of the overall quality of the essay, with no reference to specific aspects which were strong/weak. Code emerged from the data to capture praise that couldn't be coded with the criteria coding guide. 'This is a very sophisticated essay.' (Felicia) 'But overall this is really impressive!' (Celine) 'Overall, a strong essay.' (Sean)
FB8. Other-dialogic	A comment which opens a dialogue with the writer, or summarises the writer's perspective without offering a problem, solution, value judgement. 'Is it an example of Orwell's writing?' (Claire) 'I interpreted the quotation very differently. I thought that Ramazani was arguing that there were certain features of post-colonial writing that was shared, across national boundaries….' (Jess)

2. Feedback coding guide: Getting feedback

This coding guide is used for feedback recipients' responses to their feedback. This includes the following data sources: annotations made on received feedback, and Workshop 3 reflections.

The guiding question for this coding is: how is the recipient responding to the feedback they are reading?

In this coding guide, segments can be coded at more than one code.

Code name	**Explanation and examples**
G1. Agree with reviewer	A comment in which the recipient communicates exclusively or mostly agreement with the reviewers' comments. 'Yeah timing issue → too right.' (Megan, in response to feedback on her conclusion). 'Helpful example of unclear writing style' (Claire) 'Yes this bothered me, totally agree.' (Rachael, in response to commented on limited critical quotation use) 'Good. Always remember to do this.' (Phoenix) Agreement can sometimes express reservations or come across as slightly begrudging acquiescence: 'okay (though I'll have to work out how to have time to reference them…)' (Amber, in response to feedback on using more critics).
G2. Disagree with reviewer	A comment in which the recipient communicates exclusively or mostly disagreement with a reviewer comment. 'They don't contradict, they are just two separate points' (Ben) 'Not sure why this is necessary if chronology is apparent' (Amber, on giving dates for texts) 'is this a bit excessive? – every paragraph feels unnecessary' (Gemma) Can sometimes include a concession to the reviewer: 'Maybe I am wrong but I thought the quote with the question fused the two' (Megan)
G3. Intended action	A comment in which the recipient defines a possible action they could take in response to feedback. 'interesting point – balance between specific textual evidence + overall argument maybe needs work' (Celine) 'I could try conveying my points over a few sentences instead of one.' (Claire) 'I needed to spend more time planning to complete tease out how to link each text' (Megan)

Code name	Explanation and examples
G4. Dialogic	Comments or annotations where the recipient opens a dialogue with the feedback giver Annotation on essay from Holly, giving feedback: 'Good ✓ - wish this was expanded upon more / earlier in the essay' Annotation from Megan, receiving feedback: '→ I wish I had thought of it a little earlier ;)'
G5. Emotive	An annotation or comment where the recipient gives some kind of emotional response to the feedback, usually in response to praise. 'Aw what an encouraging comment!' (Rachael) 'ahh cheers' (Megan) '☺' (Amber)
G6. Self-explaining	Comment where the recipient explains their process or perspective on the written product. 'fair point (I ran out of time so knew this was pretty weak) and really helpful to have some constructive points to develop it.' (Celine) 'I could also remember limited material from poem' (Megan) 'didn't get to conclude properly unfortunately!' (Rachael)
G7. Reviewer consensus	An annotation or comment where the recipient notes the reviewers' agreement. 'same point as first review' (Chloe) '* Repeated idea (key)' (Tessa) 'this point has recurred for all 3 reviewers – needs to be taken on board!' (Rachael)
G8. Conflicting reviewer views	An annotation or comment where the recipient notes the reviewers' conflicting views over an aspect of their essay. 'Reviewer 3 disagrees + likes the 'macro' ideas of Butler – both have interesting justifications – perhaps there are other points in the entry I could cut/tidy up.' (Celine) 'diff to what R1 said' (Felicia) 'line drawn to first comment by Reviewer 1 with a ?' (Abbie)
G9. Not related to a particular comment	Used for annotations where the original comment that the recipient is responded to cannot be identified. Often used where participants make a summary of the feedback at the bottom of the page. 'summary: work on signposting in introduction use more relevant critics balance assertion with caution & understanding' (Holly) 'Timings for a comprehensive conclusion + continued interrogation of the

Code name	Explanation and examples
	quotations (both mine and the q's)' (Phoenix)
G10. New information	Used for where information given in feedback seems completely new to the recipient. Only one example so far so many be deleted. 'Hadn't thought about this.' (Phoenix)

3. Feedback coding guide: Helpful/unhelpful

This coding guide was developed in autumn 2015 to capture students' perceptions about the helpfulness of feedback. To develop the guide, I first inductively annotated and then NVivo coded responses to two reflections questions, 'Which reviewer gave the most helpful comments and why?' and 'Were any of the comments unhelpful? If so, please explain'. As this was a small section of the overall dataset, single instances of codes were allowed. Once the coding guide had been applied to the workshop items and checked for consistency, it was then applied to the section of the post-intervention interview that dealt with students' perceptions of the workshops.

The guiding question for this coding was: what did students find helpful or unhelpful in their feedback? The coding guide is in two parts: a) helpful and b) unhelpful.

A. Which reviewer gave the most helpful comments and why? (Workshop 4 item & Interview 2)

Helpful - Code name	**Explanation and examples**
Helpful - suggestions, explained how	Feedback that gave suggestions, or an explanation of how, or 'corrections' rather than just pointing out problems/issues 'Furthermore, they not only highlighted issues of my essay but suggested how these could be addressed and overcome which is really useful to me' (Chloe) 'The feedback pointed to specific parts of the essay, with actual corrections, illustrated by quotations, as well as wider points' (Joe)
Helpful - how feedback written	This code encompasses both the tone of the feedback and how the feedback was formatted or presented. 'The positivity of reviewer 1's tone also helped their feedback to feel most constructive' (Abbie) [they] 'work/separate out their advice into different bullet point style notes which means it is less of just an overview' (Gemma)
Helpful - specific, precise	Refers to the helpful precision or specificity of feedback 'Reviewer 3 gave the most helpful comments – they are very specific and detailed, picking out particular moments from my essay' (Celine)
Helpful - not coded	Parts of responses to this item that don't fit into the other codes. Codes which also only had one reference were also merged into this node (content; criteria)
Helpful - constructive	Specific uses of this term. 'Reviewer 3 gave a balanced and constructive answer' (Poppy)
Helpful - agree	Use of the verb 'agree', or implied agreement '→ Topic sentences is something I have to do better, explained how I would do this well' (Megan)
Helpful - new to author	Pointed out something that was new to the author, that they were unaware of, or they hadn't tried before 'Stylistically I found it useful how reviewer two drew attention to which areas of my writing style worked well as I had always done these things subconsciously before' (Felicia)

Helpful - Code name	**Explanation and examples**
Helpful - balance	Use of the term 'balance', or similar. 'highlighting both the good things I needed to ensure I keep and some clear, identifiable areas of improvement' (Phoenix)
Helpful - explain, justify criticisms	Criticisms were either explained or visible 'Both made really clear, well-explained points' (Rachael)
Helpful - critical	Where the author particularly values the reviewer's critiques 'They didn't spend too long complimenting and got straight to the useful and constructive feedback.' (Joe)
Helpful - thorough, comprehensive	Feedback was particularly thorough or comprehensive 'Commented on style and content, moving from more general ideas (such as more critics) to more specific ones (like in North and South) to give comprehensive feedback' (Tessa)

B. Were any of the comments unhelpful? If so, please explain. (Workshop 4 item & Interview 2 data)

Unhelpful - Code name	**Explanation and examples**
Unhelpful - not specific, no example	Where the reviewer wasn't specific enough, or didn't include an example or way to improve. Also includes a code which was initially separate: not enough information to improve. 'and "this seems a bit reductive" doesn't tell me much or make it relevant to the essay. What is reductive? What do you mean by reductive? And how? → Although he/she also raised some good points.' (Phoenix) 'Reviewer 2: "muddles your argument and creates several vague, fragmented sentences" - though this may well be valid, not pointing to specific instances makes it difficult for me to use it constructively.' (Abbie)
Unhelpful - how the reviewer comes across	Where the reviewer's tone isn't appropriate, or the reviewer doesn't come across as trustworthy 'Reviewer 4's position felt like a consistently negative position & tone that focused on any limitation it could find without offering many ways to improve.' (Ben) '"This was a really interesting essay and choice of question" was a bit generic and bordering on sympathy.' (Sean) 'Comments like "very strong" and "quite intelligent" seemed a little biased and potentially less helpful. "Sake of the reader" also seemed a little condescending to me, though I appreciate the thought behind it.' (Abbie)
Unhelpful - different knowledge, ideas, interpretation	Where the recipient sees the reviewer as drawing on a different knowledge base or interpretation to the author. 'The first reviewer… commented on my argument; this is a matter of interpretation anyway, so there could be a more constructive way to comment on the essay.' (Poppy)
Unhelpful - Student vs	Where the student is limited by their experience, and their perspective may be different to that of an examiner.

Unhelpful - Code name	**Explanation and examples**
examiner perspective	'No, most were very helpful. It was perhaps less relevant to suggest I make it clear which text I am referring to as an examiner would obviously be very familiar with those I discuss, however I will still try to do this to make my essays more clear and the direction of my points clear at the outset' (Heather)
Unhelpful - Unfair, not true to this essay	Where the recipient felt the reviewer was applying criteria unfairly or making a point that wasn't true to their own essay. 'I felt that reviewer 3's comment that I 'struggled to apply it specifically to the quotation" was unfair, because I quoted from & referred to the quotation throughout my essay' (Jess)
Unhelpful - beyond criteria, unfeasible	Where the reviewer is seen to ask for something unreasonable because it is beyond the criteria, or not feasible in the time allowed. 'Reviewer three made the comment that sometimes my essay felt too fast-paced and brisk. Whilst this is definitely true, I think it is difficult to alter due to the nature of a timed essay/exam and no suggestions were made on how I could improve this' (Chloe)
Unhelpful - not critical enough	Where the recipient feels the reviewer was not critical enough. 'reviewer 3 was very nice but they gave a lot of complements [sic] + not that much advice that linked explicitly to the essay. (their feedback <u>was</u> useful at points, I just wish they were a bit <u>more</u> critical)' (Gemma)

	C1 : Analysis	C2 : Argument	C3 : Choice of authors, texts	C4 : Clarity	C5 : Comparison…	C6 : Conclusion	C7 : Consistency…	C8 : Context & period	C9 : Critics & theory
C1 : Analysis	266	37	2	14	4	8	1	23	39
C2 : Argument	37	668	9	152	8	59	58	40	99
C3 : Choice of authors, texts	2	9	92	4	0	1	2	6	5
C4 : Clarity	14	152	4	536	8	32	28	8	23
C5 : Comparison…	4	8	0	8	146	9	1	7	5
C6 : Conclusion	8	59	1	32	9	308	22	10	16
C7 : Consistency…	1	58	2	28	1	22	159	3	6
C8 : Context and period	23	40	6	8	7	10	3	232	42
C9 : Critics & theory	39	99	5	23	5	16	6	42	621
C10 : Depth & detail	16	13	1	6	5	4	5	2	9
C11 : Development	21	46	2	21	3	30	8	7	14
C12 : Evidence	13	36	0	12	2	6	6	14	27
C13 : Finishing	0	4	0	2	0	12	0	0	2
C14 : General; broad	7	9	0	3	0	4	0	0	5
C15 : Introduction	2	104	13	56	4	39	17	12	31
C16 : Knowledge…	19	42	1	7	4	5	4	32	38
C17 : Nuance; complicating	1	14	1	6	1	3	1	0	4
C18 : Original	3	7	3	2	0	1	0	2	8
C19 : Plot	18	2	0	2	1	2	0	1	4
C20 : Progresses…	1	27	0	8	1	11	6	0	6
C21 : Question	30	164	20	103	11	66	41	35	57
C22 : Quotes	44	35	2	10	6	6	2	7	70
C23 : Range…	13	18	4	4	5	4	2	6	12
C24 : Relevance	7	15	3	10	0	4	3	15	11
C25 : Signposting…	5	95	4	79	3	18	11	5	15
C26 : Specific knowledge	5	6	0	2	1	1	1	3	6
C27 : Structure	11	50	1	26	12	13	17	5	26
C28 : Style	8	9	1	28	2	6	2	2	5
C29 : Style - concise, precise	3	11	0	13	0	7	0	2	6
C30 : Technical - referencing	0	0	0	4	0	1	0	2	16
C31 : Technical accuracy	3	9	0	22	0	1	0	1	2
C32 : Terminology	2	3	0	10	0	0	0	0	1
C33 : Voice, positioning…	13	53	1	11	2	6	3	8	51
C34 : Word count, length	1	5	0	1	0	4	0	1	2

	C10 : Depth & detail	C11 : Development	C12 : Evidence	C13 : Finishing	C14 : General; broad	C15 : Introduction	C16 : Knowledge…	C17 : Nuance; complicating	C18 : Original
C1 : Analysis	16	21	13	0	7	2	19	1	3
C2 : Argument	13	46	36	4	9	104	42	14	7
C3 : Choice of authors, texts	1	2	0	0	0	13	1	1	3
C4 : Clarity	6	21	12	2	3	56	7	6	2
C5 : Comparison…	5	3	2	0	0	4	4	1	0
C6 : Conclusion	4	30	6	12	4	39	5	3	1
C7 : Consistency…	5	8	6	0	0	17	4	1	0
C8 : Context and period	2	7	14	0	0	12	32	0	2
C9 : Critics & theory	9	14	27	2	5	31	38	4	8
C10 : Depth & detail	92	1	8	0	2	3	11	1	1
C11 : Development	1	248	10	2	3	9	5	2	0
C12 : Evidence	8	10	213	0	7	3	11	2	1
C13 : Finishing	0	2	0	26	0	3	0	0	0
C14 : General; broad	2	3	7	0	68	8	4	2	0
C15 : Introduction	3	9	3	3	8	406	3	7	4
C16 : Knowledge…	11	5	11	0	4	3	231	6	12
C17 : Nuance; complicating	1	2	2	0	2	7	6	71	0
C18 : Original	1	0	1	0	0	4	12	0	36
C19 : Plot	0	0	3	0	3	0	1	1	0
C20 : Progresses…	1	4	0	1	2	12	3	0	0
C21 : Question	9	31	21	0	7	130	49	13	12
C22 : Quotes	3	15	0	0	1	4	15	2	1
C23 : Range…	25	7	12	0	5	5	9	3	0
C24 : Relevance	5	4	8	0	3	4	10	0	3
C25 : Signposting…	1	14	3	1	2	64	5	1	0
C26 : Specific knowledge	2	2	1	1	3	2	4	1	3
C27 : Structure	2	6	8	1	2	18	7	0	0
C28 : Style	0	1	5	0	3	8	6	0	1
C29 : Style - concise, precise	0	3	4	2	1	19	3	1	1
C30 : Technical - referencing	0	0	0	1	0	2	2	0	0
C31 : Technical accuracy	0	4	1	0	0	3	1	0	0
C32 : Terminology	0	2	0	0	1	3	1	0	0
C33 : Voice, positioning…	4	4	12	0	0	18	11	1	3
C34 : Word count, length	0	2	1	0	0	1	0	0	0

	C19 : Plot	C20 : Progresses…	C21 : Question	C22 : Quotes	C23 : Range; scope	C24 : Relevance	C25 : Signposting…	C26 : Specific knowledge	C27 : Structure
C1 : Analysis	18	1	30	44	13	7	5	5	11
C2 : Argument	2	27	164	35	18	15	95	6	50
C3 : Choice of authors, texts	0	0	20	2	4	3	4	0	1
C4 : Clarity	2	8	103	10	4	10	79	2	26
C5 : Comparison…	1	1	11	6	5	0	3	1	12
C6 : Conclusion	2	11	66	6	4	4	18	1	13
C7 : Consistency…	0	6	41	2	2	3	11	1	17
C8 : Context and period	1	0	35	7	6	15	5	3	5
C9 : Critics & theory	4	6	57	70	12	11	15	6	26
C10 : Depth & detail	0	1	9	3	25	5	1	2	2
C11 : Development	0	4	31	15	7	4	14	2	6
C12 : Evidence	3	0	21	0	12	8	3	1	8
C13 : Finishing	0	1	0	0	0	0	1	1	1
C14 : General; broad	3	2	7	1	5	3	2	3	2
C15 : Introduction	0	12	130	4	5	4	64	2	18
C16 : Knowledge…	1	3	49	15	9	10	5	4	7
C17 : Nuance; complicating	1	0	13	2	3	0	1	1	0
C18 : Original	0	0	12	1	0	3	0	3	0
C19 : Plot	47	0	6	5	2	2	3	0	1
C20 : Progresses…	0	48	9	0	3	2	14	0	6
C21 : Question	6	9	1130	17	22	43	78	8	34
C22 : Quotes	5	0	17	286	6	10	3	0	4
C23 : Range…	2	3	22	6	167	2	2	2	4
C24 : Relevance	2	2	43	10	2	136	9	3	3
C25 : Signposting…	3	14	78	3	2	9	355	2	20
C26 : Specific knowledge	0	0	8	0	2	3	2	41	2
C27 : Structure	1	6	34	4	4	3	20	2	190
C28 : Style	0	0	5	4	2	0	2	0	3
C29 : Style - concise, precise	0	0	7	3	4	1	6	1	2
C30 : Technical - referencing	0	0	2	10	0	2	3	0	0
C31 : Technical accuracy	0	0	3	2	1	0	9	0	1
C32 : Terminology	0	0	2	1	1	0	3	0	0
C33 : Voice, positioning…	2	0	21	9	2	3	13	2	9
C34 : Word count, length	0	1	2	2	0	0	0	2	1

	C28 : Style	C29 : Style - concise…	C30 : Tech. - referencing	C31 : Tech. accuracy	C32 : Terminology	C33 : Voice, positioning…	C34 : Word count…
C1 : Analysis	8	3	0	3	2	13	1
C2 : Argument	9	11	0	9	3	53	5
C3 : Choice of authors, texts	1	0	0	0	0	1	0
C4 : Clarity	28	13	4	22	10	11	1
C5 : Comparison…	2	0	0	0	0	2	0
C6 : Conclusion	6	7	1	1	0	6	4
C7 : Consistency…	2	0	0	0	0	3	0
C8 : Context and period	2	2	2	1	0	8	1
C9 : Critics & theory	5	6	16	2	1	51	2
C10 : Depth & detail	0	0	0	0	0	4	0
C11 : Development	1	3	0	4	2	4	2
C12 : Evidence	5	4	0	1	0	12	1
C13 : Finishing	0	2	1	0	0	0	0
C14 : General; broad	3	1	0	0	1	0	0
C15 : Introduction	8	19	2	3	3	18	1
C16 : Knowledge…	6	3	2	1	1	11	0
C17 : Nuance; complicating	0	1	0	0	0	1	0
C18 : Original	1	1	0	0	0	3	0
C19 : Plot	0	0	0	0	0	2	0
C20 : Progresses…	0	0	0	0	0	0	1
C21 : Question	5	7	2	3	2	21	2
C22 : Quotes	4	3	10	2	1	9	2
C23 : Range…	2	4	0	1	1	2	0
C24 : Relevance	0	1	2	0	0	3	0
C25 : Signposting…	2	6	3	9	3	13	0
C26 : Specific knowledge	0	1	0	0	0	2	2
C27 : Structure	3	2	0	1	0	9	1
C28 : Style	196	4	0	12	1	12	1
C29 : Style - concise, precise	4	75	1	2	2	0	0
C30 : Technical - referencing	0	1	78	1	0	0	0
C31 : Technical accuracy	12	2	1	208	1	2	0
C32 : Terminology	1	2	0	1	34	1	0
C33 : Voice, positioning…	12	0	0	2	1	184	0
C34 : Word count, length	1	0	0	0	0	0	54

Appendix N: Standards across phases of learning

Values greater than 0 are highlighted; light blue denotes Time 1 and darker blue denotes Time 2.

	Task perc. only		Task perc. and planning		Planning only		Planning and enactment		Enactment only		Enactment and new today		New today only		Self-evaluation only		Self-evaluation and other phase		Future adaption and other phase	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
C2 : Argument	1	2	0	0	4	3	1	4	7	7	1	1	0	2	16	16	25	22	0	2
C4 : Clarity	0	0	0	0	0	2	0	0	3	8	0	1	0	6	3	11	5	13	0	2
C12 : Evidence	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	3	0	0
C13 : Finishing	0	0	0	0	0	0	0	0	1	0	0	0	0	2	5	2	7	2	0	0
C17 : Nuance; complication	3	1	0	0	2	0	0	0	4	3	0	0	0	1	1	1	2	1	0	0
C18 : Original	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	5	0	0
C21 : Question	13	13	1	5	4	12	1	3	9	25	0	2	0	5	9	19	18	24	1	1
C22 : Quotes	0	2	0	0	6	4	0	1	2	1	1	0	0	4	11	10	11	14	0	1
C23 : Range; scope	3	3	0	0	2	1	0	0	0	1	0	0	0	0	8	4	14	7	2	0
C25 : Signposting…	0	0	0	0	0	0	0	0	1	4	0	1	0	5	1	8	1	11	0	3
C27 : Structure	0	0	1	1	5	5	1	0	3	1	0	0	0	5	5	3	9	4	1	0
C34 : Word count and essay length	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	6	0	8	0	1
SRL8.1: Better planning process	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	4	12	12	5	7
SRL8.8 : Knowledge-Q match	16	15	0	0	0	0	0	0	0	1	0	0	0	0	1	0	2	1	0	0
SRL8.11 : Task perc. as standards	0	2	0	1	0	2	0	1	9	17	0	0	0	0	0	0	4	2	0	0
SRL8.12 : Time	3	1	1	0	0	0	1	0	4	2	1	1	0	2	2	4	5	5	0	1

Appendix O: Extract from Tessa's think-aloud protocol at Interview 1

Question selected:

'The "post-colonial scene" occurs in an entranced suspension of history, as if the definitive historical events have preceded us, and are not now in the making' (ANNE McCLINTOCK). Discuss.

Extract from transcript where Tessa seems to opt for this item:

If we look at the post-colonial, where are we going to go with post-colonial? 'An entranced suspension of history', what does it mean to suspend history, is this irrelevant anymore? 'It is as if the definitive historical events have *preceded* us… moving on, and are not now in the making' [writing] So the difference between the post-colonial and the colonial, suspension of history, Bhabha, and third space theory, and then, Said, Orientalism, I suppose Bhabha is a development of Said, Said seems to be more of a dichotomy between East and West. Whereas Bhabha, Third Space, mixing, the ground is not as clear. 'Occurs in an entranced suspension of history, as if the definitive historical events have *preceded* us', hmm. The texts I can refer, [writing] Village in the Jungle, Passage to India, umm… post-colonial, and then I suppose other ideas which relate to Hurston? Maybe, maybe not. Umm. Post-colonial, I suppose you've got Ireland possibly, Dubliners. 'Suspension of history, preceded us, and are not now in the making', so 'occurs in the suspension of history', so not, looking at history. Hmm. 'As if the definitive historical events have preceded us.' Definitive. 'Discuss.' Discuss, I can take that where I want, discuss. 'Suspension of disbelief, not now in the making', what do they mean by that? 'Now. In the making'. Suppose Village in the Jungle, Woolf seemed pretty ambivalent, but then there's the complexity, and idea of language barriers, mis-translation, but then I suppose you've got the clear idea of the Third Space in A Passage to India and Mrs Moore, when she comes into the mosque with Aziz, seems to understand him better, so, 'postcolonial, entranced suspension of history, preceded', [re-reads]. Hmm. But he's still colonised, they're both still colonised, 'the definitive historical events have preceded us and are not now in the making.' Looking back, the *importance* of history… [writing]. [Hums] Post-colonial, but I've only colonial texts. Theory of post-colonialism. When you're looking at it. Hmm. This is... [re-reads question quietly]

[Stops verbalising; starts writing]

Word count: 330 words.

Key:

Grey highlighting: re-reading the question: 109 words

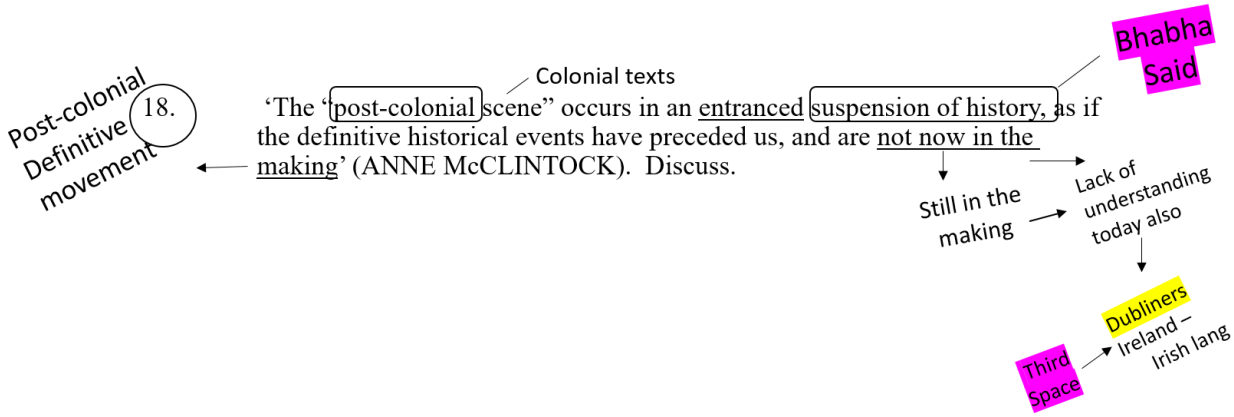
Underlining: interpreting the question: 50 words

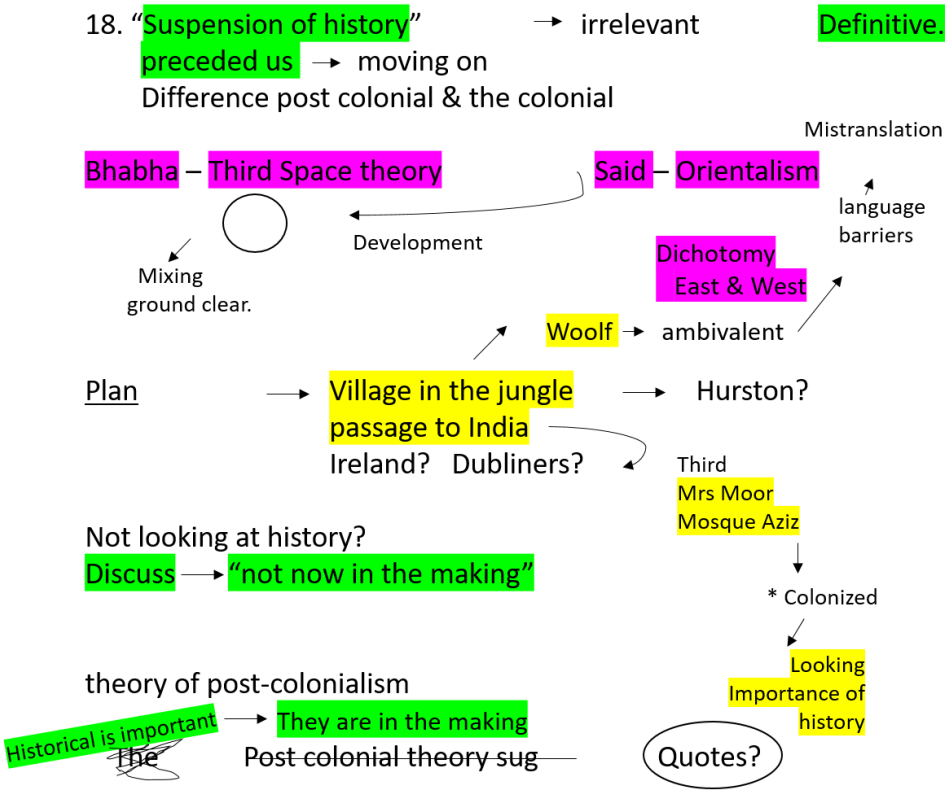
Appendix P: Tessa's planning

Planning at Interview 1 (pre-workshops)

The top figure shows Tessa's annotations on the question, the lower her written planning, on the same page where she starts writing. The final line, struck through, is a false start to her introduction.

Key: green highlighting: terms from the question. Pink highlighting: included in introduction. Yellow highlighting: included later in essay. No highlighting: not identified in essay.

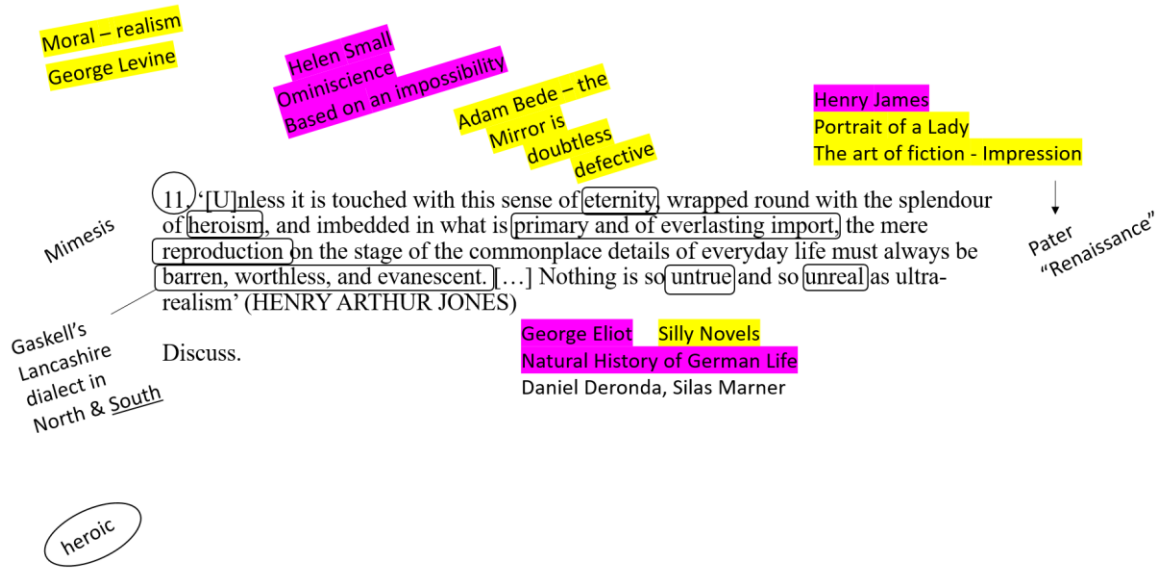




Tessa's planning: at Workshop 2 essay

The top figure shows Tessa's annotations on the question, the lower her written planning, on the same page where she starts writing.

Key: green highlighting: terms from the question. Pink highlighting: included in introduction. Yellow highlighting: included later in essay. No highlighting: not identified in essay.



Victorian 2012

Q11. "Nothing is so untrue and so unreal as Ultra-realism" → Mimesis

"mere reproduction"
"worthless"
↓
Doesn't teach anything

used as a didactic tool

"unless eternity & heroism"
Moral-realism
Ordinary events made extraordinary.

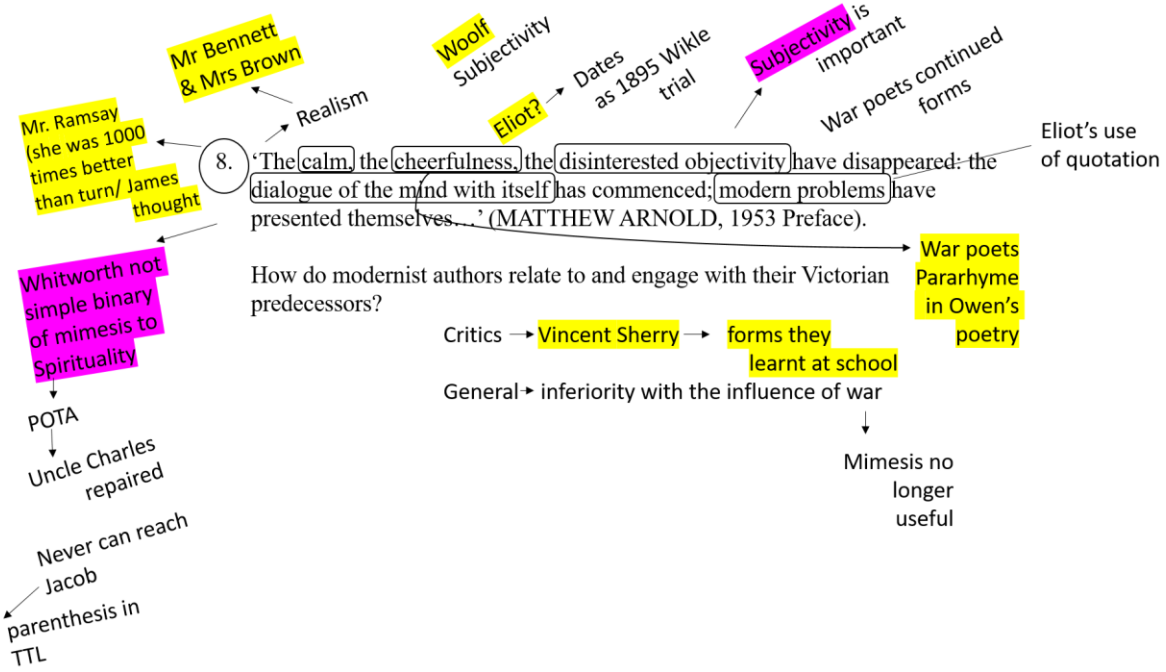
Hints at Subjectivity & perspective (Silas)

1. Moral realism (Levene) Didactic tool. Why it fails
2. Attempt to Show Subjectivity but still controlled by the author, makes sense according to a metanarrative
3. Dismisses Silly → Show life as it really is Ordinary people & their life. George Henry Lewes: realism vs falsism.

Tessa's planning at Interview 2 (post-workshops)

The top figure shows Tessa's annotations on the question, the lower her written planning, on the same page where she starts writing.

Key: green highlighting: terms from the question. Pink highlighting: included in introduction. Yellow highlighting: included later in essay. No highlighting: not identified in essay.



Plan – Wolf Mr B & Mrs B →1910 → post impressionist
Argument → Calm, cheerfulness, disinterested objective
x C19th mimesis → 20th C inferiority
with Freud
↙ relates to this idea of realism &
impressionism from Eliot
'it is before all to make you see" (Conrad)

Appendix Q: Tessa's Introductions

Key: quoting the question; ideas and keywords in plan

Essay 1: Pre-workshops – February (118 words)
'The "post-colonial scene" occurs in an entranced suspension of history, as if the definitive historical events have preceded us, and are not now in the making' (ANNE McCLINTOCK). Discuss.

(Paper 4: 1910 – present day)

"Post-Colonial" critics such as Edward Said and Homi Bhabha have examined the effects of Colonialism on English Literature and its representation of colonialized peoples. However, whilst Edward Said discusses a dichotomy between the East and West (mirroring McClintock's binary between the past and the present), Bhabha introduces the idea of a 'Third Space Theory' in which characters are presented as hybrids of the cultures and histories that surround them. In this sense, the "post–colonial scene", and realm of literary discourse need not "suspend" colonial history to focus on current events, nor purely focus on events that have occurred beforehand, but rather, incorporate them both in order to gain a better understanding of character in a post-colonial age.

critics/theory

Essay 2: During workshops – May (136 words)
'[U]nless it is touched with this sense of eternity, wrapped round with the splendour of heroism, and embedded in what is primary and of everlasting import, the mere reproduction on the stage of the commonplace details of everyday life must always be barren, worthless, and evanescent. […] Nothing is so untrue and so unreal as ultra-realism' (HENRY ARTHUR JONES).
Discuss.

(Paper 3: 1832 – 1910)

As George Eliot asserted in her essay 'The Natural History of German life' "art is the nearest thing to life", suggesting that art should be mimetic and reflect society. As Arthur Jones contests, however, "the mere reproduction on the stage of the commonplace details of everyday life must always be barren, worthless and evanescent", implying that realism as an art form is duel [sic], and to some extents "untrue". In this essay I wish to explore how George Eliot, Henry James and Elizabeth Gaskell incorporate realist techniques in their works and use them to promote moral didacticism as well as emphasising the subjectivity of perspective, in order to prove that realism in fiction (despite being surrounded on a principle of omniscience that is impossible according to Helen Small) cannot be dismissed as simply "untrue and unreal".

Author-as-critic

Critic

Essay 3: Post-workshops – June (153 words)
'The calm, the cheerfulness, the disinterested objectivity have disappeared: the dialogue of the mind with itself has commenced; modern problems have presented themselves…' (MATTHEW ARNOLD, 1853 Preface).
How do modernist authors relate to and engage with their Victorian predecessors?

(Paper 4: 1910 – present day)

The modernist focus on 'the dialogue of the mind with itself' signifies the movement away from Nineteenth century mimesis and the concern for objectivity in literature, in favour of a more subjective perspective. This alludes to the

modernist rejection of Victorian values, placing greater emphasis on the mind of the individual, especially in the wake of Freud's exploration of the unconscious in his seminal work 'The Interpretation of Dreams', translated into English in 1913. Nevertheless, as Michael Whitworth notes, 'the simplistic binary between nineteenth century omniscience and mimesis and the modernist focus on perspective and spirituality is insufficient', as twentieth century authors actively engaged with and incorporated the techniques of their predecessors. Accordingly, I wish to reexamine the relationship between modernist authors and their Victorian predecessors, drawing on the works of Virginia Woolf, and the war poets to illustrate the ways in which Modernist 'subjectivity' is indebted to Victorian mimesis and literary impressionism.

theory

critic

Commentary:

Tessa's introductions follow similar patterns, with some developments. These are discussed below, with contextualisation from her college collections essays as relevant[1].

All three introductions use ideas and keywords stated in her annotations on the question and written plan. Many of these are references to critics and/or theorists, and for Essay 2 and Essay 3, she also directly states which authors she will discuss. This is a move which she makes in one of three January collections essays and one of three April collections essays.

In all three essays, she connects with the question early on in the introduction, but the ordering of her moves varies. In Essay 1 and Essay 2, she begins with a point about critics or theory and then brings in the question; at Essay 3, she approaches the question a little more directly by analysing what part of it 'signifies', then connecting this to the literary context and Freud. In both Essays 2 and 3, she quotes a critic at some length.

In Essay 2 and Essay 3, the final sentence of her introduction states her intentions for the essay in the first person: 'I wish to explore how…' and 'I wish to reexamine how'. This move is first seen in two of three of her April collections essays, in which she uses the phrase 'I wish to demonstrate' in both introductions. As with the final sentence of the introduction in Essay 1, both statements quote the terms of the question, but her argument is clearer. The thesis statement is more direct in Essay 3, at 41 words (compared with 67 words at Essay 2).

Although Tessa notes in Workshop 1 that 'whilst the introduction should make the argument clear it does not need to be very detailed', her introductions increase in length, from 118 words pre-workshops to 153 words post-workshops.

[1] Collections are examinations set by college tutors, and usually timed for the very beginning of the term to test consolidation of the previous term's work. Students prepare for collections over the vacation. Collections are usually formative in principle, although students making poor progress may be set 'crucial' collections, the outcome of which determine whether they are allowed to continue with the course.