An investigation of experienced EAP teachers' marking processes

Author: Lijun Yin

University of St Andrews

British Council ELT Master's Dissertation Awards 2019

Commendation

An investigation of experienced EAP teachers' marking processes

Date Submitted: 17 August 2018

Table of Contents

List of Tables and Figure

Acknowledgements

Abstract

Assessment is of central importance in education in that it not only serves the purpose of certifying student achievement but also has the potential of promoting student learning (e.g. Carless, 2007). Research on assessment in tertiary contexts (e.g. Bloxham et al., 2011; Bloxham et al., 2016; Tomas, 2014) illustrates the complexity of marking as a professional judgment and problematises the transparency of marking processes in spite of the availability of assessment criteria. The elusiveness of marking criteria (e.g. Bloxham et al., 2011) and the intangibility of tutor expectations limit the learning purpose of assessment. As a response, this study reports on an investigation into experienced EAP instructors' coursework essay marking processes, an area awaiting to be explored despite the importance of writing instruction in EAP programmes (e.g. Alexander et al., 2008), with the aim of identifying what factors play in EAP instructors' perception of writing quality. 5 EAP teachers from 2 UK universities were asked to mark an essay sample and interviewed about their opinions of that essay sample and their general marking practices. The research highlights the effect of the construct of stance, as revealed in the thesis statement, developed through critical evaluation of sources and presented in a coherent manner, on the perception of writing quality, possibly because of its interrelation with explicit criteria. In addition, the application of assessment criteria is also affected by personal marking standards and student profile. Pedagogical implications of the findings of this study for the EAP instruction are also discussed.

Chapter 1 Introduction

Assessment plays an inherent part in student learning. The judgements that teachers make about student work carry high stakes for students in that they form the basis of certification for learning and impact on students' future prospects (Sadler, 2009). Scholarship on Higher Education (HE) assessment (e.g. Sadler 2009; Bloxham et al., 2011; Bloxham et al., 2016) suggests that despite the availability of assessment criteria, marking processes may still not be transparent to students, since marking, by nature, is subjective and a professional judgment. How a piece of writing is assessed involves many factors, including teachers' interpretation of criteria publicised in the institutions (Tomas, 2014), expectation of student writing which may incorporate criteria not listed in the written criteria document (Sadler, 2009; Bloxham et al., 2011; Bloxham et al., 2016;) and student profile (Bloxham et al., 2011; Tomas, 2014). Considering that criteria and writing constructs are often open to interpretation, students are likely to find difficulty in understanding teachers' expectations and feel confused with the feedback (e.g. Lea and Street, 1998; Carless, 2006) wrapped in evaluative discourses such as "be more critical" and "your essay lacks a structure". Under this circumstance, students may not know how to act on teachers' feedback, and hence the learning purpose of assessment (Carless, 2007) is undermined. Against this background, the current research investigates how EAP student coursework essays are assessed, in hopes of revealing tutor expectations and thus shedding light on the EAP instruction. This introduction first sets the context of study and then provides working definitions of key terms. The outline of this dissertation is provided at the end of this chapter.

1.1 The context of research

The internationalisation of higher education attracts an increasing number of students who use English as a second language (L2) to come to study in universities where English is the medium of instruction. With the linguistic and cultural diversity of student background, English for Academic Purposes (EAP), defined broadly as "teaching English with the aim of assisting learners' study or research in that language" (p. 1) by Hyland (2006), has now already expanded to be a major force in English language teaching. Various formats of EAP programmes are available today and cater to the needs of students at different study levels, ranging from one-year long foundation courses for pre-university students to a few weeks of summer pre-sessional programmes for postgraduates. According to Schmitt and Hamp-Lyons (2015), many EAP programmes tend to have a dual focus: one is to improve the language proficiency of learners, the other is to help students to close any gaps between their previous educational experiences and the new academic community culture that they wish to join. Hyland (2018) in his most recent publication indicates that the latter target goes far beyond than the grasp of basic language competence, and the communicative demands imposed by target disciplinary community on students can become formidable. In the light of the prestige of written communication as constructing, presenting and communicating knowledge in tertiary education (Hyland, 2013a), academic writing dominates the format of assessment in university courses and becomes the centre of EAP teaching and learning (Alexander et al., 2008). The importance of developing L2 students' ability in academic written communication in EAP programmes inspired me to look at the experienced EAP teachers' expectation of academic discourse features in student essays so as to better address student writing needs.

To make the findings of this research more generalisable, this study recruited 5 eligible EAP teachers working at two different HE institutions. Both universities provide a variety of EAP programmes including year-around EAP foundation programmes, summer EAP courses, and in-sessional language service. The 2 research sites differ in the scale, with one located in a big city enrolling the total number of students more than 4 times higher than does the other one based in a small town.

1.2 Working definitions of key terminology

In the assessment literature, various terms, which may essentially mean the same thing, are used in different occasions. To clarify the terms used in this study, the following paragraphs give working definitions of key terminology.

Assessment refers to "the process of forming a judgement about the quality of student work and extent of student achievement or performance" (p. 177), according to Sadler (2005). While formulating the judgement, teachers are also likely to provide feedback on student performance to facilitate learning, which is the key characteristic of formative assessment (Sadler, 1998). Given that evaluating and improving student learning are both the main purposes of assessment (Carless, 2007), assessment in this research is used in broad terms including the processes of scoring student work and providing feedback. Assess and mark are also used interchangeably in this dissertation to refer to such processes.

Assessment/marking criteria, also termed "rubric," "marking scheme", "marking scale", refers to "a tool used for marking student performance in assessment tasks and to provide feedback" (University of Wollongong Australia, 2015). Typically, marking criteria, as a scoring guide published to teachers and students, contain levels of performance (such as merit, high merit, distinction), cut-off point for each level (such as scores ranging from 16.5-

20 on a 20-point scale fall into the level of distinction), level descriptors (such as the key characteristics for each level of performance). In this study, criteria specifically refer to the explicit written criteria presented in a document available to teachers and students, unless modified by the adjective "implicit".

Criterion-referenced assessment /criteria-based grading means that "each student is judged against predetermined criteria without regard to other students" (Lok et al., 2016, p. 450). By contrast, norm-referencing refers to that "a predetermined percentage of students would obtain a certain grade" (Lok et al., 2016, p. 450) and often involves the comparison of student performances (Sadler, 2005).

1.3 The outline of this dissertation

The rest of the dissertation is arranged as follows. In chapter 2, literature on both HE assessment in general and academic writing assessment in particular is reviewed, and the research question, emerged from the gap in existent literature, is raised at the end of the chapter. Chapter 3 provides a rationale for the use of interviews as the data collection method and presents how data are analysed. The analysis of interview results is provided in Chapter 4. Chapter 5 discusses the findings of the current study with reference to the previous literature and suggests the implications for writing instruction and assessment. Chapter 6 summaries the key findings and points out caveats as well as directions for future research.

Chapter 2 Literature Review

As pointed out in chapter 1, this study sets out to probe into experienced EAP instructors' assessment practices of student coursework essays. Informed by the research context of academic writing assessment in tertiary education, this literature review starts with presenting the theoretical rationale of ethical considerations and educational benefits underpinning the current criteria-based assessment framework in HE contexts. What follows is an examination of empirical studies on criteria-based grading, pointing to the complexity of marking. The second part of this chapter presents literature on academic writing assessment including the necessity of stance taking in written communication and the issues in effective feedback provision arising from the elusiveness of marking criteria and evaluative discourses. After reviewing the two research tracks of educational assessment in tertiary context and academic writing, this chapter provides a summary of literature bringing together the scholarship on the two fields. This chapter ends with identifying the gap in the existing literature and raising the research question.

2.1 Criteria-based assessment in the HE context

2.1.1 Theoretical Rationale for criteria-based assessment in the HE context

In the UK higher education sector, informing students of the assessment criteria that will be used to mark their assignment is regarded as a condition for good assessment practice and thus attached much importance to by various stakeholders. For example, UK Quality Assurance Agency for Higher Education (QAA) proposes that "publicising and using clear assessment criteria are key factors in assuring that marking is carried out fairly and consistently" (QAA, 2006, p.17). In the National Student Survey questionnaire, which is issued by Office for Students (OfS) with the aim of seeking students' opinions of academic quality, the category *Assessment and feedback* contains an item named "the criteria used in

marking have been made clear in advance" (OfS, 2017). In a similar vein, students themselves also request for "clear assessment criteria linked to learning outcomes and easily accessible", as revealed in the document "Assessment and feedback benchmarking tool" published by National Union of Students (NUS, 2015). All these emphases on clear marking criteria made available to students before every assessment episode imply that assessment based on pre-determined criteria is strongly encouraged to become an established practice in HE institutions.

Theoretically, measuring students' performances against pre-set criteria is endorsed as well. According to Sadler (2005), criteria-based assessment reflects an ethical ideal, that is, students deserve to be assessed on the sole quality of their work, regardless of how well other students perform the same or equivalent tasks. On the contrary, assigning grades according to in which position a student ranks in relation to other classmates (Sadler,2005)- termed norm-referencing, can cause unfairness since individual students cannot control the performances of other group members (Sadler,2009). Moreover, norm-referencing, characterised by pre-determining a proportion of students that would receive a certain grade (Lok et al., 2016), does not align with the current outcomes-based approach to education (Spady, 1994). According to Spady (1994), outcome-based education (OBE) embraces the philosophy of "success for all students" and objects to using quotas on the number of students who can succeed. Under the guidance of OBE philosophy, the educational deficiency of successful quotas underpinning norm-referencing can potentially be overcome by criteria-based assessment which focuses on identifying the performance level with reference to criteria established on desired learning outcomes (Jawson and Askell-Williams, 2007).

Apart from translating the OBE principle into practice, criteria-based grading also has various other educational benefits in theory. In a broad educational view, Lok et al. (2016) argue that the incorporation of assessment criteria involving higher-order cognitive abilities such as critical analysis ensures that these skills receive adequate attention in teaching and learning activities. Narrowing down to individual courses, Sadler (2005) reasons that at the start of a course, notifying students of the criteria by which their work will be appraised ideally has a prospective purpose, which is to help them to refine their work intelligently before they submit. The retrospective function of criteria, according to Sadler (2005), lies in the potential of providing an explanation for assessment results reported to students, which adds more transparency to the marking process (Sadler, 2009). In a word, educational effectiveness and ethical considerations form solid theoretical foundations for criteria-based grading.

2.1.2 Empirical investigations into criteria-based grading

Despite the sound theoretical underpinnings of criteria-based grading (Sadler, 2005) and high values invested in criteria, several research on marking practices (e.g. Bloxham et al., 2011; Bloxham et al., 2016; Shay, 2005; Tomas 2014) challenges the orthodoxy of criteria-based assessment framework and points to the complexity of marking processes. In a study conducted by Bloxham et al. (2011), 12 lecturers from subjects in social sciences and humanities were asked to mark 2 essays while verbalising their thoughts. Through the analysis of the think-aloud data, Bloxham et al.'s study (2011) found that assessment criteria played a post-hoc role in refining participants' initial judgments from the grade category into an exact score or rationalising the holistic grading decisions, instead of guiding their marking processes. In addition, the use of norm-referencing is evidenced in Bloxham et al.'s research (2011), as two thirds of academic tutors either compared the 2 essay samples or

referred to a larger group of student essays in order to better capture the essay quality and monitor the scoring decisions. The disjunction between official policies and actual marking practices in higher education, whether exemplified in the post-judgement use of assessment criteria or norm-referencing, can be attributed to the reality that marking is a professional judgement involving socially constructed "personal standards frameworks" (p.668), according to Bloxham et al. (2011).

However, the claim that "a high proportion of the tutors did not make use of written criteria in their marking" (Bloxham et al., 2011, p. 655), interpreted from the observation that academics in their study did not refer to any physical assessment artefacts, needs to be treated with caution. Without distributing a set of criteria to each research participant, it is plausible that tutors were using certain criteria while marking the 2 essays but did not need to look at them, considering that they may have already internalised the criteria that they frequently use. Furthermore, the experimental nature of research of Bloxham et al. (2011) may not fully reflect the actual marking practice which has to take account of institutional requirements.

In order to address the above limitations and explore authentic assessing procedures, Tomas (2014) interviewed 12 university teachers on how they mark coursework essays. Half of the participants in her study reported that in the marking preparation stage, they would rehearse the marking criteria and gauge the standards of responses by marking a few scripts which will be re-marked later. During marking student coursework essays, tutors in Tomas's research (2014) are likely to compare student works to verify the marks that they have already given. At an individual script level, teacher markers under the study of Tomas (2014)

tend to pay attention to the global quality of student essays against criteria, along with taking a note of key points for themselves or annotating to students.

In the research conducted by Tomas (2014), participants' rehearsal of marking criteria at the beginning of assessment episodes tentatively supports that marking standards, by which judgments are formulated, may be partly shaped by stated criteria in actual marking practices, contrary to Bloxham et al.'s (2011) suggestion that teachers may ignore criteria. Nevertheless, both studies testify to the use of norm-referencing during marking processes and the focus on the overall quality of student productions in criteria-based grading. The empirical research examined above adds supporting evidence to several of Sadler's (2009) observations, which were extracted from conversations with a multitude of university teachers. According to Sadler (2009), teachers in general, instead of scrutinising student performance criterion by criterion, develop a sense of the overall quality with simultaneously taking notice of salient strengths and weaknesses, similar to the marking behaviour reported in Tomas's study (2014). Furthermore, Sadler (2009) observed that assessors often experienced disparity between holistic appraisal and evaluation of performance on the individual criteria. One of the explanations for why certain works leave an overall good impression on markers but are not rated outstanding if dissected by marking criteria, in Sadler's view (2009), is that explicit written criteria may not reflect the full range of criteria that teachers draw upon whether consciously or unconsciously. The accountability of marking criteria in the criterion-referenced assessment hence becomes problematic and calls the underlying theoretical framework into question.

More recently, Bloxham et al.'s study (2016) on how history lecturers evaluate student essays evidences the use of implicit criteria among historians, which may result in marking

variations. In addition, marking inconsistency may result from different interpretations of explicit criteria such as *developing argument* among historians in the aforementioned study, yet how assessors differ in their understandings of marking remains unexplored. To gain further insight into university teachers' evidence-based essay marking practices, research on academic writing assessment needs to be referred to.

2.2 The assessment of writing in the HE context
2.2.1 The expectation of stance in student writing
In the field of academic writing, a number of researchers (e.g. Hyland, 2008, 2012; Lancaster,2014, 2016) show a keen interest in how writers present their argument in a persuasive way to other members of their disciplinary community through stance features. In Hyland's term (2008), stance "concerns the ways writers convey their judgements, opinions and commitments (p. 7)" in order to meet the community-embedded rhetorical expectations (Hyland, 2012). Through a corpus analysis of stance features in 240 academic papers from leading journal articles in 8 disciplines, Hyland (2008) developed a model of stance elements, comprised of hedges, boosters, attitude markers and self-mention. His model shows that expert academic writers are likely to use hedging devices to withhold full engagement with a statement. At the opposite end, boosters are also taken on by authors in Hyland's sample (2008) to express their conviction in an argument. Attitude markers signalled by words such as *interesting*, and the explicit author reference including first-person pronouns are the remaining two stance patterns in Hyland's stance model (2008). Derived from the analysis of discourse features in expert writing, Hyland's (2008) theoretical framework of writer-oriented stance and reader-oriented engagement, which are covered under the umbrella term "voice", is influential in academic discourse studies and informs later research on voice and stance.

Filtering down to student writing, stance features are, though maybe implicitly, rewarded by teachers. Based in the US HE context, Lancaster (2016) carried out a comparative study of stance, including the four stance options in the Hyland's model (2008) reported above, in addition to the fifth device called disclaim marker such as *but, however*, in high- and low-graded undergraduate essays from 2 disciplines. Across the 2 courses, student papers with a high grade show a significantly higher number of stance elements in total than the corresponding low-level writing. In terms of the five categories examined, the frequency of the appearances of hedges and disclaim markers in high-marked coursework essays is statistically significantly higher than those linguistic devices in low-marked ones in each course. The extensive use of hedging expression by high-achieving first-year students in Lancaster's (2016) corpus-based investigation is consistent with the top research articles examined in Hyland (2008), in which hedges are the most commonly used stance feature. As for disclaim marker, Lancaster (2016) argued that it is used for problematising other people's views and helps to index a critical stance. Although the expectation of critical thinking was articulated by professors from both courses in interviews, Lancaster (2014, 2016) pointed out that language features that can be used to realise such stance may run below the faculty staff's full awareness.

However, the predominantly quantitative method used in Lancaster's study (2016) fails to provide a holistic picture of how stance affects the writing quality perceived by teacher markers without any qualitative assessments involved. Moreover, the frequency counts of stance features conducted by Lancaster (2016) might, though perhaps inadvertently, suggest to readers that a more frequent use of stance features would result in a better stance expression and a higher essay quality. A positive correlation between the

effectiveness of stance delivery and the number of stance elements is not entirely agreed upon by participants in Zhao's (2012) empirical study on how to reliably measure the strength of voice, or in Lancaster's term "stance" (2016), encompassing both writer-oriented stance defined by Hyland (2008) and writer's stance towards the potential reader.

Zhao's (2012) study, mentioned above, is a mixed-method research on the assessment of stance positioned in the context of high-stakes writing exam. With the use of both think-aloud method and statistical analysis, Zhao (2012) identified and empirically validated the three dimensions of stance, which are the clarity of ideas, the manner of idea presentation, and writer and reader presence. The involvement of human raters in Zhao's (2012) research highlights that the linguistic elements raised in Hyland's (2008) model of writing interaction can contribute to effective stance projection only if they are used properly. In addition, the presence and clarity of a central point, which are a global discourse feature not addressed by Hyland (2008), were valued by participants in Zhao's study (2012) in their assessment of stance realisation. However, whether the positive effect of the articulation of a central point on the stance quality can be generalised to source-based coursework essays is not clear, since Zhao's study (2012) was based in the assessment setting of TOEFL independent writing test.

In terms of the coursework essay genre, the construct of stance delivered globally with the development of argument was investigated by Wingate (2012) through the analysis of tutor comments on the first-year university students' writing. Viewing stance from a broad angle, Wingate (2012) proposed that writer position and stance need to be established upon a critical evaluation of sources and presented in a coherent manner. With reference to his theoretical framework, Wingate (2012) found that the failure to construct a stance valued

by teachers results from ineffective source use, including lack of criticality, lack of evidence, unrelated information and lack of structure. However, tutors' expectation of stance may not be well communicated to students, as Wingate (2012) surprisingly discovered that position was not explicitly mentioned in tutor comments under study. The absence of the word of "position" or "stance" in Wingate's study (2012) may align with Lancaster's (2016) statement that stance is "an important though hidden" (p. 16) feature, perhaps buried under other explicit criteria such as structure and argument and impressionistically assessed (Zhao, 2012) by tutors. The importance and expectation of stance expression, though recognised by literature on academic discourse studies, may not be fully realised by students and teachers, if stance is not included in the assessment criteria.

2.2.2 Issues in providing effective feedback to students

Apart from evaluating student performance as indicated in a grade, teachers tend to be also concerned about helping students improve their learning in an assessment event. The learning purpose of assessment (Carless, 2007) is an important area that has been widely researched on through the study of teacher feedback. According to Nicol and Macfarlane-Dick (2006), good feedback practice has at least the characteristics of clarifying what good performance is, notifying students of their own performance level and providing opportunities to close the gap between present performance and intended learning outcomes. For feedback to deliver its learning potential, Carless (2007) further adds that student engagement with feedback is the key aspect in addition to its timeliness. A starting point for students to effectively use feedback, as argued by Carless (2006), is to understand the evaluative discourses underlying the feedback. The comprehensibility of feedback tends to be examined from students' point of view, sometimes in comparison with tutor perceptions, by researchers (e.g. Carless, 2006; Wingate, 2012).

Previous studies on how tutor written comments are interpreted by students through interviews (e.g. Lee and Street, 1998; Carless, 2006; Bloxham and Campbell, 2010) and diary studies (e.g. Wingate, 2012), however, uncovered the difficulty in understanding tutor expectations delivered through feedback. For instance, a first-year university student in Lea and Street (1998) showed confusion over the comment that "your essay lacks structure" despite his attempt to achieve cohesiveness through using connective words and referring back to the key terms of the title. Critical thinking, another important writing construct, is also a potential cause of miscommunication in Wingate's study (2012) which compared tutor comments with student diaries. The new university entrants studied by Wingate (2012) reacted in their diary that they felt too inferior to criticise famous academic figures in their subject area when they were told that they lacked critical awareness in the feedback. By contrast, criticality for tutors in Wingate's study (2012) may mean the synthesis of sources, as evidenced from the comment on high-graded essays that sources are effectively synthesised. The problem in unpacking assessment discourses including both criteria and feedback is echoed by Carless (2007) and Bloxham and Campbell (2010) based in a Hong Kong university and a British HE institution, respectively. Students' lack of understanding of what is expected of them in academic writing may partly explain why improvements are often not made with advice from feedback, which took teachers considerable time to produce (Sadler, 2010).

2.3 Summary of literature

In the current outcomes-based approach to higher education, evaluating student performance against criteria established on intended learning outcomes has the theoretical advantages of adding transparency to marking processes and helping students make improvements in their work. However, empirical studies on assessment (e.g. Bloxham et al.,

2012; Bloxham et al., 2016; Tomas, 2014) point to the complexity and subjectivity of teacher grading. Instead of processing individual criteria discretely, markers are more concerned about how a piece of work comes together as a whole (e.g. Sadler, 2009; Bloxham et al., 2012; Tomas, 2014) with decisions on levels of performance in terms of individual criteria made retrospectively (Sadler, 2009). Bloxham et al. (2012) concludes that marking is in fact a professional judgement involving tutors' tacit knowledge, interrelated criteria and socially constructed standards. To further explore how the overall writing quality is perceived, research in academic written discourses needs to be drawn upon.

Studies on academic writing (e.g. Hyland, 2008; Lancaster, 2014) illustrate the importance of expressing stance in a way that is accepted by other members of certain disciplinary communities in achieving successful written communication. Investigation into how the construct of stance is assessed empirically (e.g. Zhao, 2012) or through the analysis of tutor comments (e.g. Wingate, 2012) suggests that stance is related how ideas are presented and argument is developed. Despite of its value in writing quality, stance is likely to be a feature hidden from faculty staff's awareness (Lancaster, 2014, 2016) and impressionistically assessed (Zhao,2012). The implicit construct of stance and the intangibility of explicit criteria such as critical analysis (e.g. Wingate, 2012), may not only cause student difficulty in understanding what is expected of their writing but also taking feedback advice into action.

2.4 Gap in the literature and research question
The evaluation of the quality students' written responses, whether approached from educational assessment in tertiary contexts (e.g. Bloxham et al., 2012, 2016; Tomas, 2014; Sadler, 2009) or academic writing (e.g. Lancaster, 2014; Wingate, 2012), is mainly analysed through the viewpoints of academics. A search of *Journal of English for Academic Purposes* and *Assessment and Evaluation in Higher education* shows that research attention has not

yet been paid to EAP instructors' marking processes of students' evidence-based coursework essays. Given the central role of writing in EAP instruction (Alexander et al., 2008), it is believed that an enquiry into how student writing is assessed by EAP teachers in the safe research environment with anonymity assured can make contributions to EAP teaching and learning. With reference to literature, EAP teachers can be made more aware of their expectations of student work and better identify and address student writing problems. To fill the gap in the EAP assessment literature in the HE context the present study is conducted in order to answer the question "how do experienced EAP teachers mark students' coursework essays?"

Chapter 3 Methodology

Interviews are carried out in order to answer the research question about how EAP teachers mark student coursework essays. In this chapter, the participants of this study are introduced first, and then information about the two research instruments is presented. What follows is the account of how research data are obtained and analysed. This chapter ends with the ethical considerations.

3.1 Participants

Five EAP teachers, 4 females and 1 male, from 2 high-ranking universities in Scotland participated in this study. All of the participants are British national citizens with English as their first language. Every participant has rich experience in teaching EAP courses ranging from 14 years to 22 years. The EAP programmes that they are involved in and/or have taught cover a wide range of variety including pre-sessional summer EAP courses for postgraduates, foundation programmes for pre-university students or 1st year undergraduates, EAP programmes for pre-masters and in-sessional language service.

3.2 Instruments

3.2.1 Essay sample

The around 1500-word essay sample (Appendix A) used in this study was adapted from a website of a university in England. The essay is a written response to the prompt "Countries that intervene in trade have cited political arguments for their actions. To what extent are these arguments justifiable? Discuss with reference to examples and rationales." The assessment task was done for an undergraduate module titled "global business environment". The piece of work, updated in November2016 and published by the learning centre in the School of Management of the aforementioned university, is annotated and

originally given to 62%. A look at the annotation implies that the essay sample may be flawed in introduction, paragraphing, source use and referencing. The open access, imperfectness, and proper length make this essay an appropriate choice for this research.

However, in order to avoid any confirmation bias, all comments in the essay sample were removed, and only the module information is retained. The original mark was not provided to participants either. Given that the contextual information provided in this study is rather limited, participants' marking of the essay may not fully represent their daily marking practices when all kinds of information such as learning outcomes and classroom input are available. However, asking every participant to mark the essay sample was designed to elicit EAP tutors' expectations of student writing and compare their perceptions of the writing quality.

3.2.2 Marking criteria

All participants in the study were distributed to the same marking criteria (Appendix B) adapted from the criteria used in one of the research universities. The criteria are divided into 2 columns: levels and level descriptors. There are 5 levels in total, including fail, pass, merit, high merit, and distinction. For each level, there is a list of key characteristics on the corresponding right column. However, participants are not obliged to stick to the one grid for all individual criteria; they can highlight the level of individual criterion in different boxes if they feel that the work reaches different levels for different criteria. The 7 criteria categories extracted from the level descriptors are: knowledge and understanding, language and vocabulary, critical awareness, source use, structure and argument, approach to the task, clarity and cohesion. The cut-off grade in the marking criteria was eliminated, given that the criteria adapted from the one research site are based on a 20-point scale, while percentage marking is used in the other research university. In order to reduce the effect of

a familiarity with the marking scheme on the sample essay evaluation, participants were not asked to give a grade to the essay sample but highlight the descriptions on the criteria that they felt the essay matched and comment on the essay sample.

3.3 Data collection and analysis procedures

3.3.1 Research design and pilot of study

The aim of uncovering EAP teachers' marking processes determines that this research is by nature a predominantly qualitative study, the characteristic of which involves rich descriptions (Mackey and Gass, 2015). Among the qualitative approaches to data collection, interviews were used since they allowed researchers to investigate events, in this case marking processes, that were "not directly observable" (Mackey and Gass, 2015, p.225). More specifically, semi-structured interviews with a list of pre-set questions as a guide (Mackey and Gass, 2015) are the preferred method. The rationale for selecting this type of interview is that it makes comparisons among participants possible by asking the pre-determined questions meanwhile leaving space to probe for more information.

With the research plan in mind, the semi-structured interview questions were designed and piloted on an EAP teacher, who was not a participant in this study. The data elicited from the pilot study were transcribed and analysed, and interview questions were found effective in guiding speakers to talk about their marking practices. Given that there was no essay sample used in the pilot, abstract questions such as "how do you understand critical awareness", though answered by the participant in detail, were not described with examples. In order to help EAP teachers clarify their points, the original research design was improved. Specifically, an essay sample was sent to all the participants to comment on with the same assessment criteria. The use of the essay sample and criteria was also helpful to compare how assessment criteria were applied by different teachers.

The renewed research design was piloted on another EAP teaching staff who was, again, not a participant of the study. The time spent on the second-round pilot helped me to inform the participants about the estimated time, which was about 45 minutes, so that they can arrange appointments with sufficient time length. Meanwhile, the semi-structured interview questions were again found helpful to elicit useful data addressing the research question. After a minor revision and 2 rounds of piloting, the semi-structured interview protocol was settled and provided in appendix C for readers' reference.

3.3.2 Teacher marking and interview
The research design, including teacher marking and interviews, was communicated to the potential participants through emails, together with the participant information sheet and the coded data consent form. After confirming to take part in the study, participants were sent the essay sample and marking criteria a few days in advance so that they can have enough time to mark. They were told to highlight the level of criteria that they felt the essay sample reached. During the interviews, participants were first asked about their background information. After that, they were interviewed about their opinions of the essay sample and then asked questions related to their general marking practices. It is noticeable that when asked about their marking work in general, participants were likely to refer to the essay sample to illustrate their points, which further supported the current research design. At the end of interviews, teachers were asked to send back the criteria that they had highlighted and the annotations that they had made, if any, to supplement the analysis of interview data.

3.3.3 Coding of qualitative data
After the interviews were completed, the audio-recorded data were non-verbatim transcribed. The focus on the content of each participants' speech made it safe to ignore the

words such as uh, well that did not contribute to the meaning for the purpose of saving time on transcribing. A sample of transcription is provided in appendix D for readers' reference.

With all the interviews transcribed, the coding work started with reference to Saldaña (2009). Each transcript was read carefully for several times for developing a familiarity with research data and went through two cycles of coding. The structural coding method, which resulted in identifying segments of text on broad topics (MacQueen et al., 2008, as cited in Saldaña,2009), was applied to analysing individual transcripts in the first cycle of coding. Specifically, guided by interview questions and marking criteria used in the study, the key points of each participant's answers were summarised and categorised into topics. After coding an individual interview transcript, it was read again to check that the coded extracts were representative of the participant' answers on each topic. As all transcripts went through such procedure, the first cycle of coding ended.

The codes and topics from individual transcripts were brought together for further thematic analysis (Saldaña,2009) in the second cycle of coding. Under each broad topic, the codes from the first cycle were read and compared, and sorted into themes. Each theme was checked and refined with regard to the coded extracts and original interview transcripts. By the end of the second cycle coding, the similarities and differences among participants' opinions on certain topics became clear.

In addition, the participants' ratings of the level of the essay quality on each individual criterion were also tabulated to get an overall pattern of marking consistency and variation. The next chapter will present the data analysis results in detail.

3.4 Ethics

The current research was approved by the school ethics committee. After obtaining the

school ethics approval, each potential participant was sent an invitation email attached with

the participant information sheet and the coded data consent form at least a week before

the interview started. They were told to feel free to ask any questions before they decided

to participate. For example, 2 participants raised their concerns about the item "I agree for

my tape-recorded material to be used in future studies" listed in the coded data consent

form format provided by my school. They were reassured that the data would be only used

by me at any time. With all questions answered, participants signed the consent form and

agreed to be audio-recorded on the day of the interview. The 5 participants were

anonymised using teacher A to E throughout this dissertation.

Chapter 4 Results

This chapter, divided into the two main parts, provides an analysis of data obtained from interviews. The first section compares the teachers' evaluations of the essay sample. The participants' overall ratings of the level of the essay on individual criteria are first provided, followed by a more detailed analysis of teachers' qualitative description of the essay quality. The major strengths and weaknesses of the essay sample perceived by each teacher are summarised, and common themes in the essay evaluation as well as marking differences are identified and interpreted.

The next section focuses on participants' general marking practices, which contains both grading and feedback. The analysis of participants' scoring-decision making processes is further divided into three sections: reading strategy, application of assessment criteria and the role of student profile in marking. Participants' practices in providing both written corrective feedback and summative feedback are analysed at the end of this chapter.

4.1 Evaluation of the essay sample
4.1.1 Ratings of the level of essay on assessment criteria
Generally speaking, the essay sample falls into the level of merit or high merit by given criteria, except *critical awareness,* to which all participants assigned merit, and *clarity, cohesion (and concision),* with one participant highlighting the distinction for this category. In terms of individual criteria, *language & vocabulary* was achieved at a satisfactory level with 3 participants giving high merit, followed by *structure & argument* and *approach to the task* each thought to be at the high merit level by 2 participants. The greatest variation in participants' evaluation of the essay sample was found in *clarity, cohesion (& concision)* with 3 merits, 1 high merit and 1 distinction. *Knowledge & understanding* and *source use* are the

two criteria that were highlighted at the level of high merit by only one participant E and B respectively. The ratings of the essay on individual criteria by every participant are presented in Table 1.

A comparison across the marking results produced by 5 participants found that teacher B and E from the two different universities gave high merits or above for most of the criteria. Teacher A and D also of different institutional affiliations show exactly the same judgements of the writing quality by all criteria while teacher C highlighted all the key characteristics at the merit level. The overall rating pattern formed the basis for classifying the 5 participants into 3 groups in the following analysis: teacher B and E with overall higher rating, teacher A and D of the same rating, and teacher C sticking to the merit regarding each criterion.

Criterion	Teacher A	Teacher B	Teacher C	Teacher D	Teacher E
Knowledge & understanding	Merit	Merit	Merit	Merit	High merit
Language & vocabulary	High merit	Merit	Merit	High merit	High merit
Critical awareness	Merit	Merit	Merit	Merit	Merit
Source use	Merit	High merit	Merit	Merit	Merit
Structure & argument	Merit	High merit	Merit	Merit	High merit
Approach to the task	Merit	High merit	Merit	Merit	High merit
Clarity and cohesion (& concision) *	Merit	High merit	Merit	Merit	Distinction

Table 1   Participants' ratings of the essay on each criterion

Note * Concision is included in the high merit and distinction level descriptors while the statement in the merit level is "reasonable clarity and cohesion."

4.1.2 Perceptions of writing quality
4.1.2.1 Teacher B and E
Judging the essay high merit or above by most of the given criteria, Teacher E thought that the writing sample "was very good", and she would give it overall "between merit and high merit". The good writing performance was attributed to her strong perception that apart from the very good essay structure, "language was very good in general, and certainly the

argument flowed. Certainly this person knew a lot about this subject and was able to express it."

Reflected in the level of merit for only 2 criteria categories which are *critical awareness* and *source use*, lack of criticality and too few sources were recognised as the two main problems of the essay sample. According to teacher E, the unquestioning acceptance of "a few beliefs and attitudes" demonstrated the essay writer's lack of criticality. A specific example is the paraphrasing sentence that "despite some most countries claiming to be committed to free trade, they need to intervene in order to attain political benefits (Daniel et al. 2009: 305)". Her remark for this citation is that "it's not they need to intervene; it's they **think** they need to intervene," revealing the need for questioning literature. On the other hand, claims with no "evidences to back them up" signal too few sources. Taking note of the reference list, Teacher E further commented that the use of newspapers was not expected in an academic essay; rather, "more peer-reviewed journal articles" were encouraged. Despite the deficiency in the selection and use of sources, the overall impression is still that the essay is written at a high level.

For teacher B, who gave 4 high merits on the assessment criteria, thought the essay was "okay" and generally "dealt the question quite well". To him, the strength of the essay lies in the arguments which are "pretty good", "quite strong" and "answer the question". Despite some short paragraphs in the essay, which were not a major issue for him, he gave overall highly favourable comments in terms of paragraphing: "There's nothing wasted. All the points attempt to go back to the question, and the examples are related to the question." The weakness of the essay structure, however, lies in the introduction since it lacks a central

statement and contextual information. Figure 1 is the snapshot of his comment on it, which is also representative of opinions of the rest of the participants except teacher E.

Countries that intervene in trade have cited political arguments for their actions. This essay will explore some of these arguments, and assess the extent to which these arguments are justifiable. Each argument will be presented and discussed with reference to examples and an analysis of the motivations of countries.

This introduction could be stronger. More context is required at the beginning to set the scene for the question and there is no indication of an overriding opinion to guide the answer – no thesis statement, no clear outline of points to be covered.

Figure 1   Teacher B's comment on the essay introduction

Another weakness that drew his attention was the limited range of expression. Most notably, phrases such as "it could be said", "many would claim" and "it would be argued" are "just repeated," which negatively affects the style of the essay and probably reduces the language of the essay to the merit level rated by teacher B, in contrast to the high merit level of language perceived by teacher E. However, the weakness in the introduction pointed out by teacher B but not shared by teacher E did not cause any differences in their ratings of the essay in the aspect of *structure and argument* since they both gave high merit on this criterion. It seems that teacher B, as he claimed, gave credits for the essay author's commitment to the topic, the evidence of which as stance expression will be further discussed in the next chapter.

4.1.2.2 Teacher A and D
Similar to teacher B, teacher A and D, who both thought that the essay reached the level of high merit in terms of *language and vocabulary* but merit on the other criteria, also commented on structural flaws, especially in the introduction and paragraphing. For teacher D, faced with the current "very short intro", she advised the student to follow the "BITS (Background-issue-thesis-scope) model of writing an introduction." Considering the prompt

"to what extent", teacher D argued that such question asked for "a strong stance", the lack of which can be addressed by providing "a clear thesis statement."

In teacher A's opinion, the "poor introduction" failed to provide her "sufficient guidance", "took her a while into the essay" and did not "meet [her] expectation at all". When asked further what her expectation is in terms of introduction, teacher A's answer reveals the similar spirit as teacher D's BITS model: "a general statement, contextual definitions, rationale, overview of the argument, problem to be discussed, and then a strong thesis statement that shows the writer's position". However, according to teacher A, the ineffectiveness of introduction in framing the essay argument was partly compensated by "a helpful conclusion." Constructed in "more of a standard" way, the conclusion gave "an overview of the arguments and "fill[ed] some gaps that [she] had remaining from the beginning."

From "a poor introduction" to "a helpful conclusion", teacher A noticed a developmental pattern in terms of paragraphing. To her, the "undeveloped and unconventionally structured" 3 paragraphs at the beginning made it hard to decipher what was being communicated, but later paragraphs were "pretty good in terms of cohesion and developing links between ideas." By the end of the essay, she drew the conclusion that "there [was] a coherence to the argument", which means the argument was "logical" and on the topic, but the essay lacked "cohesiveness" because of the flawed beginning paragraphs.

This upward trend in the essay structure recognised by teacher A, however, was not agreed upon by Teacher D. From the latter's perspective, the lack of "sophistication" in terms of paragraphing, despite "some very good examples", can be improved by combining "two paragraphs talking about the same example" and joining two examples on the opposite

points of view "together in one paragraph". Unlike teacher A's focus on the problems of beginning paragraphs, teacher D's suggestions of combining paragraphs are distributed throughout the essay.

The language aspect of the essay was lauded by both teacher A and D and thought to fit into the high merit level descriptors, although the two teachers differed in the impact of the misuse of punctuations on the argument. Teacher A was of a favourable opinion that the language was "excellent" with "a couple of minor punctuation errors" and occasional "incidences of informality". In comparison, teacher D remarked that "the student ma[de] a lot of mistakes by overusing punctuation", especially in the use of "semi-colons", and punctuation problems "definitely need[ed] to be addressed" since it "[did] impact on the flow in places". Still, the essay left the teacher D an overall impression of "very good language", due to the "very good grammar", "mainly appropriate style of language" and "enough variety of language". It is worth noting that the last point, though mildly commended by teacher D, received warning from the aforementioned teacher B, who recommended that the writer pay attention to language variation, and even stronger criticism from teacher C discussed below.

4.1.2.3 Teacher C
Compared with all the other teachers who more or less found strengths in the essay, teacher C's comments are predominantly negative. Overall, the essay is "quite poorly written" and has "a very poor style" even though it "has not got many language errors" in individual sentences. "The main problem" of the essay, however, lies in the "lack of cohesion", which "spoils all the critical evaluation", and the use of "weak examples", evidenced by teacher C's questioning of the essay author's referring to Iran and Cuba, where situations "are totally different", for justifying US interventions. In addition, the use

of sources is also unsatisfactory for teacher C, demonstrated by her comment on one quotation that "needs to be integrated into the text" and suggestion of "making your own point first" on a paragraph starting with a citation.

Structurally, the "extremely poor introduction" with "no thesis", according to teacher C, was "useless" since it "[didn't] tell [her] what stance that the writer was taking." While this limitation was echoed by most of participants, teacher C, in sharp contrast with teacher A who thought the ending paragraph was helpful, instead found the conclusion equally disappointing. Her criticism was that it merely "summarise[d] the main points" without "really answering the question" and to make matters even worse, "[gave] another example which it really shouldn't be doing". The length of body paragraphs is also problematic in teacher C's viewpoint, as she felt that "a lot of paragraphs need[ed] to be put together properly.

As for language, teacher C pointed out the misuse of semi-colons, which was "quite disturbing to the reader". Apart from the repetitive use of "an example of" which seemed annoying, the overuse of hedging phrases such as "perhaps", "maybe", "it can be said" significantly affected the delivery of a clear stance in teacher C's opinion. The following examples show her critical evaluation of the improper use of tentative expressions:

"To start with a paragraph with '**it could be argued that**'. What does that tell me? It tells me that they don't really know what they're trying to argue".
"**perhaps it is not justifiable**.' By the end of paragraph, why **perhaps**? They are just scared of saying anything".
 The above comments illustrated that hedging devices, used unskilfully, left teacher C the impression of no stance in the essay, the implication of which will be further discussed in section 5.4 of the next chapter.

4.1.2.4 Summary of participants' perceptions of the essay sample

Table 2 summarises the five teachers' opinions of major strengths and weaknesses of the essay sample emerged from both interviews and written comments

	TEACHER A	TEACHER B	TEACHER C	TEACHER D	TEACHER E
Major Strengths	Very good language Coherence to the argument Effective paragraphs Helpful conclusion	Strong argument Ideas and examples related to the topic Generally accurate grammar	Not too many language errors	Very good language -very good grammar -formal style -enough variety Very good examples	Very good language Cohesion of the argument Very good Subject knowledge
Major Weaknesses	Poor Introduction Citations needed to be integrated Lack of cohesiveness in the beginning paragraphs	Poor introduction Lack of language variety	Poor style Lack of cohesion between sentences Poor use of sources and examples Structure -Useless introduction and conclusion -short paragraphs punctuation errors too much hedging	Poor introduction Paragraphs needed to be combined Lack of cohesion between paragraphs Punctuation errors	Lack of criticality Too few sources

Table 2   Strengths and weaknesses of the essay sample

It can be seen from table 2 that grammatical accuracy was generally appreciated, although the extent to which this feature affected the judgement of overall language quality differed among teacher assessors, since other aspects of language such as variety and punctuation also came into play. The defects in the essay, on the other hand, seemed to gain consensus from most of the teachers under study in certain aspects such as ineffective introduction,

which was mentioned by 4 participants and led to the perception of lacking a stance. The

lack of critical awareness in the essay sample was another commonly perceived deficiency,

which was reflected in all participants' assigning the level of merit to this criterion. How

these two factors-stance and critical awareness- were understood by participants is

discussed as follows.

4.1.3 Common themes in the essay evaluation

4.1.3.1 The need of taking a stance

The need of showing a stance was acknowledged by all participants, most notably reflected

in the expectation of seeing a clear stance in the thesis statement at the very beginning.

Even though teacher E did not mention the shortcomings of the introduction, she stated

that she looked for an "independent thought" when marking the writing sample and

"absolutely" encouraged her students to give their own opinions. The importance of stance

was strongly endorsed by teacher A and C as well:

"If the marker is not clear about your position, then you can have really well-developed
arguments but don't fall under anything...These arguments are coherent and cohesive, but
they don't connect up to the central point." (Teacher A)

"If you don't have a stance, it means you're not engaging with the topic." (Teacher C)

The above extracts testify to the negative effect of lacking a stance at discourse level.

Linguistically, teacher C believed that the overuse of hedging "[was] very inappropriate to

get in a good stance" as exemplified in the essay sample and illustrated in section 4.1.2.3.

The importance of stance expression in the perception of writing quality will be further

explicated in section 5.1 of the discussion chapter.

4.1.3.2 Shared understanding of critical awareness

Considering the importance of criticality in the essay writing (e.g. Lancaster, 2014),

participants were asked about their conceptions of this notion. A marked model, composed

of "claim—evidence—counter-evidence—refutation," emerged from their answers and may

explain why they all agreed that the essay sample only demonstrated reasonable critical

awareness. Below are the example quotations of participants' responses to the question

"how do you understand critical awareness?"

"Are they providing a claim? Are they explaining the claim, giving an example, then giving a counterpoint to that, then refuting that?" (Teacher A)

"Critical awareness is how they [students] synthesise different authors' viewpoints. Put forward your argument, and you counter, and you back again, say however" (Teacher B)

"What you're doing is that you're saying this person has convinced me by his argumentation more than that person's convinced" (Teacher E)

As shown above, critical awareness is closely related to the criterion source use and

underlies stance expression. The interaction of the construct of stance with other criteria

will be discussed with reference to literature in the next chapter.

4.1.4 inter-rater differences in the essay evaluation
4.1.4.1 Interpretation of criteria
Despite the similar view of critical thinking, divergences in the interpretations of other

criteria are identified among participants. As language teachers, their judgements of the

essay language indicate the sub-criteria underlying the category *language and vocabulary*

may result in rating differences, or in some cases, lead to the same marking results but for

varying reasons. Table 3 (see next page) exhibits the comments on and ratings of *language*

and vocabulary of the essay sample.

As observed from table 3, punctuation errors were noticed by teacher A, C and D. While

teacher A considered that they were just minor mistakes, teacher C and D felt a detrimental

effect of the misuse of semi-colons on the fluency of argument. For teacher D, this negative

impact may be offset by other positive language features since the language of the essay

was still regarded to reach the level of high merit. However, the variety of language,

regarded as sufficient by teacher D, is the main limitation with regard to the criterion *language and vocabulary* in the viewpoint of teacher B, which may reduce the level of language quality to merit instead of high merit. From teacher C's perspective, "good language [in the case of this essay] would have been reducing the hedging and so on"; she nevertheless regarded that the essay fell into the merit level descriptor in terms of *language and vocabulary*.

	Comment on the *language and vocabulary* of the essay	Rating of the level of *language and vocabulary* to the essay sample
Teacher A	Excellent language Minor punctuation errors Very occasional incidences of informality	High merit
Teacher B	Generally accurate grammar Limited range of expression	Merit
Teacher C	Too much hedging language The misuse of semi-colons disturbing to the reader Lack of links between sentences even though individual sentences are perfectly okay	Merit
Teacher D	Very good language Very good grammar Appropriate style Enough language variety Punctuation errors that impact the flow of argument	High merit
Teacher E	Very good language Minor grammar mistakes	High merit

Table 3   Participants' comments on and ratings of *language and vocabulary*

The inter-rater disagreement masked by the same rating is manifested from the appraisal of the *structure and argument* of the writing sample too. In spite of the summative descriptor *clear structure and argument* highlighted by both teacher A and C, they have completely different attitudes towards the conclusion with the former praising it "helpful" and the latter criticising it "useless". Whether leading to marking variation or not, the difference in the interpretation of criteria can be attributed to different personal standards, which affirms the findings of previous literature as shown in section 5.1.2 of the next chapter.

4.1.4.2 The impact of the perception of stance on marking differences
Except for variance in the interpretation of given criteria, markers' perceptions of the implicit construct of stance were likely to play a role in assessing the piece of writing that they were sent to, though participants themselves may not be aware of it. For example, teacher B attached importance to "whether somebody answers the question or not" in his marking of the essay sample in specific and student essays in general. "Writing around the topic" was believed to be a common issue among students, and the writing sample in question was credited for it "certainly answering the question." This overarching concern of students' commitment to the topic, which in turn is a sign of stance realisation, tends to dominate his interpretation of the given criteria and therefore rating of the essay, especially in terms of *structure and argument.* Next chapter will further discuss the interaction of the implicit construct of stance and explicit criteria.

On the other hand, teacher E articulated that the first thing she looked for was an "independent thought" when marking the essay sample; however, instead of letting this construct embedded in her use of written criteria, she seemed to regard it as a separate category and "look[ed] for good and solid performance at each level of criteria," if she did not find the author's original thought. She further explained that it was the lack of individual opinions that prevented the essay from standing out and getting distinction, despite her overall rating was slightly higher than other teachers except *critical awareness* and *source use*. The marking variation may point to the differing extent to which the implicit construct of stance, in teacher E's words, "independent thought" and explicit criteria interact, a representation of the idiosyncrasy of personal standards.

4.2 General marking practices of coursework essays

4.2.1 Scoring decision-making process

4.2.1.1 Reading strategy

When reading and assessing EAP student essay scripts, most of the participants in this study suggested that they looked at the overall structure first and then went through to see how arguments were developed, except teacher C who addresses the scoring task at a group level. Specifically, teacher C would "read all of scripts and correct all and only language errors", then "have a rest and re-mark each script".

At an individual script level, teacher A's reading strategy is slightly different from other teachers. To "get an overall sense" of writing quality, she first "look[s] at the introduction and conclusion" and then "pick[s]a paragraph" to see how it is developed, followed by "skimming and picking up language features". Teacher B, D and E, instead, would "read the whole essay first" to get an overview of the macro-organisation of the piece of writing. The reading strategy adopted by participants may suggest that they initially focus on the global quality of student works, which is in alignment with research literature as discussed in chapter 5.

4.2.1.2 The application of assessment criteria in essay scoring

Most of the participants believe that they "mark to the criteria (Teacher C)", whether in the form of broad dimensions "language, content and structure (Teacher D)" or specific categories including "structure and organisation, argument, cohesion and coherence, language use, referencing and citations (Teacher A)." The strong adherence to criteria was particularly asserted by teacher E: "I don't always agree with the criteria, but if I were presented with a set of criteria, I have to mark an essay according to that criteria." Starting from "the overall impression", teacher B thinks that criteria help him adjust his first impression and become more "objective" in essay scoring.

When asked whether it is difficult to apply criteria in scoring, teacher C and D attributed the easiness of matching student writing with levels on criteria to their years of experience. As an experienced EAP tutor, teacher A, however, finds it time-consuming to determine "which of the criteria descriptors fit [her] thought of the essay." With 18 years of EAP teaching experience, teacher E "often finds it difficult to mark an essay" unless they are extremely good or poor. Still, a strong adherence to criteria is perceived by all participants, but as discussed later, criteria may still not guide their marking processes.

4.2.1.3 The role of student profile in essay scoring
Despite the claim of marking to the criteria, teacher C is aware of the necessity of knowing the "overall standard" of a particular group by reading all of the scripts, otherwise there is "nothing to measure against" when marking the first one among a group of students. Similarly, teacher B recognises the lack of "context" causes the imprecise score for the first paper he marks and thus often re-marks it. Although not directly mentioned by teacher A, the building of a marking context is implied from her picking two students at the higher and lower end of performances respectively to see whether they "meet [her] expectations" if she has student knowledge through her course. What this suggests is that grading purely on assessment criteria, though theorised in the framework of criteria-based grading, may not be practical, which adds supporting evidence to the existing HE assessment literature.

4.2.2 Feedback provision
4.2.2.1 Treatment of language errors
Varying approaches to error correction are identified among teacher practitioners in this study. Viewing herself a lecturer on language courses, teacher C, who "specifically mark[s] the language as well as other things", seems very enthusiastic about error correction. She notices that an abundance of "silly language mistakes" make a potentially very good essay "difficult to read" and "put it down [from distinction] to merit", which may explain her

behaviour of correcting all grammar mistakes. For teacher E, whether to correct errors or not is subject to the nature of programmes that students are enrolled in. At the beginning of one-year long foundation courses, for example, teacher E would "mark every single grammatical mistake" as teacher C but does not continue this practice to avoid "demotivating" students. If students are in summer pre-sessional classes, she would not focus on grammatical mistakes but just remind students of particular language problems once. The nature of mistakes, on the other hand, plays a role in teacher A's error correction behaviour. She feels "very quick to pick up on" stylistic inappropriateness, but only comments on "systemic" grammar mistakes that "run through the essay" in the end instead of commenting on each one. The responsibility of pointing out salient grammatical problems in feedback is shared by teacher B and D, but language errors that they feel distracting are also commented. Nevertheless, most of the participants will notify students of the language errors that cause communication breakdown.

4.2.2.2 Techer belief in providing feedback
Although not prompted, most participants articulated their philosophies in providing summative feedback to students. While teacher B is of the view that feedback should tell "strength" and space for improvement, Teacher D "firmly believe[s]" that the best feedback is composed of three elements: what the good points are, "what needs to be improved, and specific suggestions on how to improve". In addition, she is very cautious of not changing "what the student is trying to say" in her comments, but aware that sometimes she does it "without being deliberate".

In teacher C's feedback practice, the current level of students is taken into account to ensure that they "have means to implement." The consideration of students' ability to use feedback, in teacher A's case, results in "targeted feedback" rather than "comprehensive

feedback" since "lots of detailed information is detrimental" to students' capacity of taking

advice into action. The varying strategies adopted in feedback provision, together with

whether teachers should correct student errors, are discussed in section 5.3 of the next

chapter.

Chapter 5 Discussion

This study sets out to investigate EAP instructors' marking processes through interviews. An analysis of teachers' comments on the essay sample, as presented in chapter 4, found the effect of the construct of stance on the perception of writing quality among participants. This chapter starts with the discussion of this construct in the essay assessment. The next main section 5.2 is devoted to the discussion of the application of explicit assessment criteria, followed by the discussion of feedback provision. Implications for EAP writing assessment and instruction are provided at the end of this chapter.

5.1 Discussion of writing constructs for EAP teachers in the essay assessment
5.1.1 The effect of stance on the perception of writing quality
This study found that EAP participants look for stance realised in a way that conforms to their conceptions of knowledge construction in student writing. Having worked in the EAP teaching profession for at least a decade, participants in this study have similar expectations of essay introduction represented by teacher D's BITS (background-issue-thesis-scope) model and critical awareness summarised in the formula of claim—argument—counter-argument—refutation. While using these models to evaluate essay quality consciously or subconsciously, EAP practitioners confess that they hope to find student writer's "stance" (teacher C, D) in the thesis statement (teacher A, B, C, D) or "independent thought" (Teacher E).

The requirement of students taking a stance and voicing individual opinions but meanwhile conforming to conventions can be explained if stance is looked at from a social-constructivist perspective, which takes account of the contributions of both individuality and social norms in the construction and perception of an authorial stance (Matsuda, 2015).

According to Matsuda (2015), the social-constructivist framework of writer identity seeks stance "in the perceived reality that results from the text-mediated interaction between the writer and reader" (p. 149). Markers in this study, who have established membership in their respective EAP community based in the British HE system due to their rich teaching experience, expect to see a position developed through critical evaluation of sources in students' written texts. This preferred way of stance realisation is in line with previous empirical research conducted by Wingate (2012), in which effective evaluation and synthesis of source materials characterise high-marked essays, and echoed by Hyland (2008), who argues that acknowledging alternative views in literature makes stance more credible. The absence of individual opinion, demonstrated in the unquestioned acceptance of what is said by certain scholars, prevents the essay sample from standing out and reaching the level of distinction, at least for teacher E in this study.

For the other research participants, the lack of articulating a central point at the beginning of the essay affects the display of a clear stance. Likewise, this discourse-level stance element is attached importance to by raters measuring voice strength in the study carried out by Zhao (2012) and found to be a significant predictor of essay scores obtained in high-stakes argumentative writing assessment (Zhao and Llosa, 2008). In addition to the absence of a thesis statement, the lack of coherence between sentences perceived by teacher C and paragraphs by teacher A and D points to the effect of the presentation of ideas, a major component of voice conceptualized and empirically validated by Zhao (2012), on the rating of *structure and argument* in specific and writing quality in general. In a similar vein, the deficiency in organising ideas in a logical manner is typical of low-achieving essays in Wingate's study (2012). The presence of clear ideas and presentation of ideas in a coherent manner (Zhao,2012) are therefore two dimensions of textual voice, in Hyland's term (2008),

"stance," valued by most of the EAP teachers in the study. The shared implicit conception of stance by participants reinforce Hyland's (2008) definition that stance is a "community recognised personality" (p.7).

5.1.2 Interaction of implicit construct and explicit criteria

The above analysis illustrates that stance, though not directly mentioned in the assessment criteria distributed to participants, is related to the extent to which *critical awareness* is demonstrated through *source use* holistically and impacts on the rating of *structure and argument* at discourse level. While the absence of a central point in the introduction paragraph of the essay sample captures participants' attention and is regarded as a failure to get stance across, teacher E seems not to take note of this discursive feature, which may contribute to a higher rating of *structure and argument*, despite her holistic impression that the essay lacks "independent thought."

Another problematic side of stance realisation in the writing sample perceived by some participants-the lack of cohesiveness in the presentation of ideas-is neither mentioned by the aforementioned teacher E nor teacher B, the other one who gives "high merit" to *structure and argument*. Instead, teacher B appreciates relevant ideas and examples and sees them as a sign of the author's commitment to the essay question under discussion, which in turn functions evidence of stance-taking. The different rating results due to varying interpretations and assessments of stance illustrate Bloxham et al.'s view (2011) that marking in reality is a professional judgment involving implicit knowledge, complex and interrelated criteria. It is interesting to note that the overall higher rating given by teacher B and E adds tentative evidence to the influence of the key elements of *structure and argument* in essay writing (Lee and Street, 1998; Wingate,2012) on the other criteria.

At a linguistic level, the overuse of hedging, an important constituent in Hyland's stance

model (2008), is thought by teacher C to negatively affect the establishment of a strong

stance and lead to her assigning merit to the category of *language and vocabulary.* On the

other hand, the extensive scholarship on hedging devices (e.g. Hyland, 2000, 2008) finds

that hedges frequently appear in prestigious journal articles from various disciplines. The

judicious use of this linguistic feature by expert writers helps reflect a sensitivity towards

other members of the disciplinary community and exhibit an appropriate academic persona

(Hyland, 2000), while the improper toning down the assertiveness of claims can leave the

impression that the writer has no confidence in his/her opinions and hence fails to show a

clear stance. Here again, the assessment of the construct of stance, though unstated in the

written criteria, not only exerts an influence upon the appraisal of *language and vocabulary*

of the essay sample but also causes the overall writing quality to be marked down by

teacher C.

5.2 Discussion of the application of assessment criteria in marking practices

All of the participants reported that they use predetermined criteria in their general marking

practices, although the difficulty in using criteria is still acknowledged by teacher A and E

with many years of experience in teaching and marking. The adherence to criteria, in spite

of the occurrence of discomfort with it sometimes for teacher E, is in line with the essence

of criteria-based assessment, which presumes that marking becomes more robust across

markers if a complex judgement is broken down into small-scale judgements (Sadler, 2015)

on each constitutive criterion. However, participants' approaches to marking tasks through

first "reading through" (teacher B, D, E) to "get an overall sense" (Teacher A) of writing

quality, consistent with Sadler's observation (2009), illustrate that assessors initially take an

interest in the global quality of student works other than in the performance on each

criterion. Sadler (2009) further elaborates that decisions on what levels student works reach on the individual criteria are made retrospectively, as teacher A tries to "match the criteria descriptors to [her] thoughts". Criteria in turn can refine teacher B's overall impression, the function of which is also evidenced in an empirical investigation by Bloxham et al. (2011) in which criteria were found to be used for checking a holistic decision.

In reporting and explaining appraisal outcomes, criteria are invariably resorted to by assessors given that criteria constitute evaluative discourses, according to Sadler (2015). An analysis of participants' opinions of the essay sample, together with their self-reported data about their general marking practices, illustrates that the interpretation and actual use of criteria involve "personal standards framework" (Bloxham et al, 2011, p. 665). For instance, the variety of language is viewed as "enough" by teacher D but "limited" in teacher B's views, which may be the contributing factor in their marking variation for the category of *language and vocabulary*. The difference in applying criteria may result from teachers' dissenting expectations.

Apart from teacher expectations, the level of group performance in general also plays a role in participants' application of criteria. The supply of marking "context" (Teacher B) is built through "reading all of scripts" together (Teacher C) and selecting two works of students at each "end of performance levels" (Teacher A). Being opposite to criteria-based assessment and not espoused in stated policies, comparing students' performances and norm-referencing is a recurrent theme in the marking research, examined through using the think-aloud method (Bloxham et al, 2011) and interviews (Tomas, 2014). Lok et al. (2014) summarised that most teachers, out of pragmatic concern, hope to see certain level of absolute performance meanwhile respecting conventions in grade distribution. Likewise,

EAP lecturers under this investigation demonstrate that their personal standards of marking are subject to student profile. Their behaviour of norm-referencing may be out of concern with the gate-keeping purpose of writing assessment in the university (Hyland,2013a).

5.3 Discussion of feedback provision
As teachers for ESL writers, participants in this study show varying attitudes towards providing corrective feedback as reported in section 4.2.1. Whereas teacher C explicitly "corrects all language errors", for other teachers whether to correct language errors depends on "length and stage of course" (teacher E) and "nature of mistakes" (teacher A). The ways in which errors are corrected are also divided among participants: with telling students their language problems in the summative feedback instead of "correcting every mistake" (Teacher B), indirect correction of mistakes is also found in the error correction behaviours of teacher B, D, and E.

The best practice in terms of whether and how to correct errors, however, is not defined in research literature on corrective feedback. Hyland and Hyland (2006) attribute the lack of a conclusive answer to the complexity of second language acquisition in that L2 students follow a U-shape pattern of language development, where learners' initially acquired ability in using the target form may backslide before grasping it again. The non-linear growth in language proficiency, according to Hyland and Hyland (2006), makes it highly questionable to establish a clear relationship between error correction and successful acquisition of certain language form. What is clearer is that outside EAP classrooms, corrective feedback is not likely to be provided for the purpose of developing students' language proficiency by subject tutors, since academics are more concerned about how students use the medium of writing to demonstrate their subject knowledge (Hyland and Hyland, 2006). Students themselves also realise that mistakes are tolerated by their subject lecturers, provided that

language inaccuracy does not undermine getting argument and ideas across in Hyland's study (2013b). However, his study, based in a Hong Kong university where English is the medium of instruction, also uncovers that EAP teachers' effort in improving students' English grammatical accuracy is still appreciated, as one undergraduate reported that "it's really bad to make a mistake and not know that it's a mistake. "(Hyland, 2013b, p. 180). EAP teachers therefore share "the responsibility of reminding students of their language problems" (Teacher D) if students' needs are taken into consideration, despite the recognition that "accurate grammar does not [sufficiently] make a good essay" (Teacher C).

As for feedback in general, participants adopt various strategies to ensure that students are able to "implement" (Teacher C) written comments including giving "specific suggestions" (Teacher D) in addition to pointing out strengths and weaknesses and providing "targeted feedback" (Teacher A) instead of comprehensive feedback. Giving constructive advice for students to act on and prioritising areas for improvement are in accordance with the principle of good feedback practice raised by Nicol and Macfarlane-Dick (2006). Making use of individual student's ability level in teacher C's response to student writing illustrates the view that teachers create a context for their feedback, taking account of their knowledge of the student writer and tailor their remarks to student needs (Hyland and Hyland, 2006). The use of formative assessment in EAP courses incorporating both certification of student achievement and learning element(Carless, 2007) requires that instructors "split their attention between grading and writing feedback" (Teacher A) and deliver high-quality information to students about how their current state of performance relates to standards defined by assessment criteria (Nicol and Macfarlane-Dick, 2006) and interpreted by individual teachers. In this regard, developing students' understanding of their teachers'

expectations needs to be a primary concern in feedback provision, be it in the written form, or oral conference, or a combination of both.

5.4 Implications for EAP writing assessment and instruction
The above discussion suggests that the assessment of a myriad of students' written responses is subject to many factors including how student writers realise their individual stances, how well student works reach levels of assessment criteria interpreted by EAP teachers and negotiated by group performance. In order to bridge the gap between students' present achievement and next level of criteria, teachers try to employ strategies to help students act on feedback. However, research on learners' uptake of feedback (e.g. Lea and Street, 1998; Carless, 2006; Bloxham and Campbell, 2010) evidences that novice student writers, whether they are home students or come from different cultural backgrounds, have difficulty in comprehending teachers' expectations that are wrapped up in elusive evaluative discourses, and hence may not be able to make effective use of feedback, despite the considerable time devoted to its production.

An illustrative example revealed in this study is the conception of stance, which may cause miscommunication to happen. The finding that stance, in particular, and writing quality, by extension, are related to how hedging devices are used instead of how often they appear is consistent with the perception of raters who discount the frequency-based measure of voice strength in Zhao's study (2012). Despite the scholarly notion that hedges are used with the intention of distinguishing opinion from fact (Hyland, 2008), Lea and Street's research (1998) documented that expressions such as "it can be said" may be employed by novice student writers, who have been taught that "I" is not appropriate in academic writing (Hyland, 2002), for the substitution of first personal pronouns. Hiding personal opinions behind hedging language can inadvertently leave the impression that the writer is "scared

of" (Teacher C) engaging with the topic, and hence the writing lacks a strong stance. What this implies is that these linguistic features are not strict rules to follow, but tools, if applied appropriately, can help portray a strong stance and subsequently improve writing quality (Zhao and Llosa, 2008). Although several teachers in this study are against using models for fear of dampening student creativity, exemplars such as the essay sample used in this study can help clarify the message that rules are not rigid but needs to be used strategically to achieve the intended effect.

At discourse level, the finding that stance is constructed through critical evaluation of sources can be another cause of miscommunication. The notion of *critical awareness*, which is shared among the community of EAP practitioners in this study and means a synthesis of sources from scholars of varying opinions, may not be well understood by students. For example, first-year undergraduates in Wingate's study feel that they are powerless to criticise famous academic figures because of their lack of knowledge, similar to a personal experience reported by teacher E. The mismatch in the conceptions of criteria between teachers and students is echoed by Nicol and Macfarlane, 2006), which may be a cause of poor writing performance. According to Sadler (1989), knowing what a good performance is forms a necessary condition for students to benefit from feedback. However, teaching staff in the study conducted by Bloxham and Campbell (2010) suggested that students' own "fairly unsophisticated" rules (p.295), such as the misunderstanding of critical thinking discussed above, govern how they are going to write and improve their essays. EAP instructors and students alike thus need to be conscious of the insufficiency of published criteria in developing standards that resemble established members of particular community of practice. Again, dialogue about carefully chosen exemplars (Carless and Boud, 2018) can be used to draw out students' interpretation of criteria and illuminate teachers'

assessment expectations such as how to effectively portray a stance and demonstrate critical awareness.

Not only can students' understanding of criteria differ from university tutors, teachers themselves also show variation in the application of writing constructs to evaluate student works. Compared with academics in Bloxham et al.'s study (2016), EAP teachers in the present study show less dramatic marking disparity. Still, the notion of stance seems to exert varying effects on teachers' evaluation of the essay sample, although their comments largely associate stance with a thesis statement, a narrower view compared to the literature on stance or voice (e.g. Matsuda, 2015; Zhao and Llosa, 2008; Zhao, 2012). Defined in broad terms, stance, an encompassing construct that brings together explicit criteria, discussed in the previous section 5.1, may be able to capture teacher markers' holistic impression of texts. The implication of research on stance can inform the writing instruction and assessment in a way that it may shed lights on how to improve a piece of work as a whole.

Finally, participants' attitudes towards language indicate that EAP tutors share a responsibility of drawing students' attention to their language inaccuracy and more importantly, helping L2 learners employ their linguistic resources to achieve communicative effectiveness. This dual duty requires teachers to point out L2 students' linguistic problems since research (e.g. Hyland, 2013 b) suggests that students want to know their mistakes and lay emphasis on their acquiring new discourse practices in the EAP instruction. Moreover, teachers in this study mention that their views of the essay sample are perhaps different from subject tutors. Since written communication involves producing a text that the writer presumes matches the expectations of potential readers (Hyland, 2008), EAP tutors may need to consider the extent to which the writing quality perceived by their community is

different from members of the target community and the effect of differences on students'

acquisition of academic discourses, especially when students are about to exit EAP

programmes.

Chapter 6 Conclusion

As pointed out in chapter 1, this study investigates how student coursework essays are assessed by experienced members of the EAP community. This interest not only stems out of the importance of academic writing instruction in EAP programmes (e.g. Alexander et al., 2008), but also the lack of research in the area of EAP assessment in tertiary contexts. As argued by Schmitt and Hamp-Lyons (2015), the research in EAP assessment is largely focused on standardised language proficiency tests such as TOEFL and IELTS. It is believed that bringing the processes of marking student coursework essays into conscious awareness can help clarify what kind of performance is expected. As a response, this research sets out to address the topic of essay marking starting from reviewing literature on HE education assessment in general and academic writing assessment in particular, since assessing academic writing in tertiary education is the intersection of these two research tracks.

Choosing 5 experienced EAP instructors from 2 HE institutions in the UK, this qualitative study used interviews to investigate the EAP teachers' specific marking experience of an essay sample and marking practices in general. The interview results were analysed and discussed with reference to literature.

Beginning with this brief overview of the whole research, the rest of this chapter is arranged as follows. A summary of the key findings is provided next, followed by the contribution of the current research to EAP instruction. Caveats in the present research and directions for future research are included in this chapter as well.

6.1 Summary of the key findings

The research question "how do experienced EAP teachers mark students' coursework essays?", including both grading and providing feedback, is answered from 3 main aspects.

In terms of the implicit construct of stance in academic writing, the study found that EAP teachers expect students to show their stance in a way that conforms to the community-embedded value of knowledge construction. Although not included in the provided assessment criteria, the construct of stance may still affect EAP instructors' perceptions of the overall writing quality since it interacts with other explicit criteria. The two broad dimensions of stance expression -the presence of a central point and the presentation of ideas in a coherent manner- interacts with criteria *argument and structure* and *clarity and cohesion*. Holistically, stance is expected to be established on the basis of effective synthesis of *sources*, which requires students' *critical awareness*. The employment of linguistic devices such as hedges in stance expression may affect the perception of the level of *language and vocabulary*.

The explicit marking criteria, though may be divided into discrete items as presented in this study, are rather interrelated. Instead of looking for levels of performance on individual criteria, teacher markers tend to develop a sense of the overall quality of student writing and then make decisions on individual criteria retrospectively (e.g. Bloxham et al., 2011; Sadler,2009.) The differences in the application of criteria are also evidenced in this research, perhaps because of different personal standards. In addition, the level of group performance on each assessment episode may also play a part in shaping teachers' marking standards, which is consistent with previous literature (e.g. Bloxham et al., 2016; Tomas, 2014).

In terms of feedback provision, the study evidenced that for most EAP participants, whether and what kind of corrective feedback are provided are subject to many factors such as stage of programme, nature of mistakes and students' future needs. Various strategies such as targeted feedback and providing specific advice apart from telling strengths and weaknesses are adopted by EAP instructors with the intention of helping students engage with feedback more effectively.

6.2 Contributions of my research

It is hoped that the discussion of the above findings with reference to previous research can better inform EAP writing instruction. The discrepancy between teachers' perceptions and literary discussions, however, is most noticeable in the linguistic realisation of stance. Although most participants recognise the importance of a thesis statement in the stance expression, the linguistic aspect of stance features, though widely researched on (Hyland, 2000, 2008; Lancaster,2016) is only directly mentioned by one participant. An awareness of the impact of linguistic devices, on stance projection can help teachers give advice on students' language choices.

In addition, teachers need to be aware that sometimes their expectations and requirements such as critical thinking, though shared in their community, may not be well communicated to students. The misunderstanding of what good performance is and tutor expectations is even more likely to happen to students who are from distinct cultural backgrounds. For instance, Alexander et al. (2008) argues that Chinese rhetoric traditions assume that the reader has a role in filling the background knowledge and understanding the implied meanings of written texts, in shape contrast with the explicitness of academic writing in English. The difference in academic cultures may cause such student group to have a wrong

perception of desirable performance and thus produce an essay which may match their own notion of good academic writing but not their teachers'. Facing students from different cultures and educational backgrounds, EAP teaching staff should pay attention to making their expectations more explicit, for example, through dialogue about carefully chosen essay exemplars, so that students' understandings of academic discourses align with the one valued by the target disciplinary community.

6.3 Caveats

This study is limited in several ways. Despite the advantages in open access, appropriate length and imperfectness, the essay sample used in this study was perceived to be written by a native English speaker by all participants. Combined with the lack of contextual information, this specific marking experience may not fully represent their actual marking work, where the majority of students use English as a second language. The limited sample size with predominantly female teachers and the use of only one essay sample in this study may also affect the generalisability of the research findings. Nevertheless, the selection of participants from two different institutional contexts may help to make the study findings more transferable.

6.4 Directions for future research

As for future researchers who are interested in this area, the improvements can be made on research design. A mixed-methods approach, combining both tutor qualitative assessments and linguistic analysis of more essay samples, may be better able to capture the writing quality. The difficulty in finding explanations for why teachers have different perceptions for the same criteria experienced in this study may be addressed by using the alternative qualitative data collection method such as think-aloud protocols.

References

Alexander, O., Argent, O., and Spencer, J. (2008) *EAP essentials: A teachers' guide to principles and practice*. Garnet Publishing Ltd.

Bloxham, S. and Campbell, L. (2010) 'Generating dialogue in assessment feedback: Exploring the use of interactive cover sheets', *Assessment and Evaluation in Higher Education*, 35(3), pp. 291–300.

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(6), pp. 655–670.

Bloxham, S., den-Outer, B., Hudson, J., and Price, M. (2016) 'Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria', *Assessment and Evaluation in Higher Education*, 41(3), pp. 466–481.

Carless, D. (2006) 'Differing perceptions in the feedback process', *Studies in Higher Education*, 31(2), pp. 219–233.

Carless, D. (2007) 'Learning-oriented assessment: Conceptual bases and practical implications', *Innovations in Education and Teaching International*, 44(1), pp. 57–66.

Carless, D. and Boud, D. (2018) The development of student feedback literacy: Enabling student uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8), pp.1315-1325.

Hyland, K. (2000) 'It might be suggested that…': Academic hedging and student writing. *Australian Review of Applied Linguistics*, 16, 83–97.

Hyland, K. (2006) *English for academic purposes: An advanced resource book*. Routledge.

Hyland, K. (2008) 'Disciplinary voices: Interactions in research writing', *English Text Construction*, 1(1), pp. 5–22.

Hyland, K. and Hyland, F. (2006) Feedback on second language students' writing. *Language Teaching*, 39, pp.83-101.

Hyland, K. (2012) Undergraduate Understandings: Stance and voice in final year reports. In K. Hyland and C. Guinda (eds.) *Stance and Voice in written academic genres.* Palgrave: Macmillan.

Hyland, K. (2013a) Writing in the university: education, knowledge, and reputation. *Language Teaching,* 46(1), pp. 53-70.

Hyland, K. (2013b) 'Faculty feedback: Perceptions and practices in L2 disciplinary writing', *Journal of Second Language Writing*, 22(3), pp. 240–253.

Hyland, K. (2018) 'First Person Singular Sympathy for the devil? A defence of EAP', *Language Teaching,* 51(3), pp. 383–399.

Jawson, M. and Askell-Williams, H. (2007) *Outcomes-based education*. Available at http://www.eduhk.hk/obl/files/pratical_guide_5.pdf.

Lancaster, Z. (2014) 'Making Stance Explicit for Second Language Writers in the Disciplines: What Faculty Need to Know About the Language of Stance-Taking', *WAC and Second-Language Writers: Research Towards Linguistically and Culturally Inclusive Programs and Practices*, pp. 269–298. Available at: http://wac.colostate.edu/books/l2/chapter11.pdf.

Lancaster, Z. (2016) 'Expressing stance in undergraduate writing: Discipline-specific and general qualities', *Journal of English for Academic Purposes*, 23, pp. 16–30.

Lea, M. R. and Street, B. V. (1998) 'Student Writing in Higher Education: An academic literacies approach', *Studies in Higher Education*, 23(2), pp. 157–172.

Lok, B., McNaught, C. and Young, K. (2016) 'Criterion-referenced and norm-referenced assessments: compatibility and complementarity', *Assessment and Evaluation in Higher Education*, 41(3), pp. 450–465.

Mackay, A and Gass, S. (2015) *Second language research: Methodology and Design.* Routledge.

Matsuda, P. K. (2015) 'Identity in written discourse', *Annual Review of Applied Linguistics*, 35, pp. 140–159.

National Union of Students. (2015) *Assessment and feedback benchmarking tool.* pp.1-2. Available at: https://www.plymouth.ac.uk/uploads/production/document/path/2/2533/NUS_Assessmen A_and_feedback_benchmarking_tool.pdf.

Nicol, D. and MacFarlane-Dick, D. (2006) 'Formative assessment and selfregulated learning: A model and seven principles of good feedback practice', Studies in Higher Education, 31(2), pp. 199–218.

Office for students. (2017) *National Student Survey 2017-Core Questionnaire.* pp 1-2. Available at: https://www.thestudentsurvey.com/content/NSS2017_Core_Questionnaire.pdf.

Quality Assurance Agency for Higher Education. (2006) *Code of practice for the assurance of academic quality and standards in higher education. Section 6: Assessment of students.* pp. 1-44. Available at: http://dera.ioe.ac.uk/9713/2/COP_AOS.pdf.

University of Bradford. (2016) *Sample Essay-UG Global Business Module, with ELS comments*. Available at https://www.brad.ac.uk/management/media/management/els/Global-Business-Environment-sampler-essay.pdf. Accessed on 16 June 2018.

University of Wollongong Australia. (2015) *Assessment Glossary.* Available at https://www.uow.edu.au/curriculum-transformation/aqc/assessglossary/index.html. Accessed on 16 August 2018.

Sadler, D. R. (1998) 'Formative Assessment: revisiting the territory', *Assessment in Education: Principles, Policy & Practice*, 5(1), pp. 77–84.

Sadler, D. R. (2005) 'Interpretations of criteria-based assessment and grading in higher education', *Assessment and Evaluation in Higher Education*, 30(2), pp. 175–194.

Sadler, D. R. (2009) 'Indeterminacy in the use of preset criteria for assessment and grading', *Assessment and Evaluation in Higher Education*, 34(2), pp. 159–179.

Saldaña, J. (2009) The Coding manual for qualitative researchers. Sage Publications Inc.

Schmitt, D. and Hamp-Lyons, L. (2015) 'The need for EAP teacher knowledge in assessment', *Journal of English for Academic Purposes*, 18, pp. 3–8.

Shay, S. (2005) "The Assessment of Complex Tasks: A Double Reading." *Studies in Higher Education*, 30 (6), pp. 663–679.

Spady, G. (1994) *Outcomes-based education: Critical issues and answers.* Available at https://files.eric.ed.gov/fulltext/ED380910.pdf.

Tomas, C. (2014) 'Marking and feedback provision on essay-based coursework: a process perspective', *Assessment and Evaluation in Higher Education*. Routledge, 39(5), pp. 611–624. doi: 10.1080/02602938.2013.860078.

Wingate, U. (2012) '"Argument!" helping students understand what essay writing is about', *Journal of English for Academic Purposes*. Elsevier Ltd, 11(2), pp. 145–154. doi: 10.1016/j.jeap.2011.11.001.

Zhao, C. G. and Llosa, L. (2008) 'Voice in high-stakes L1 academic writing assessment: Implications for L2 writing instruction', *Assessing Writing*, 13(3), pp. 153–170.

Zhao, C. G. (2012) 'Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric', *Language Testing*, 30(2), pp. 201–230. doi: 10.1177/0265532212456965.

Appendix A Essay sample

Level: Undergraduate

Global Business Environment Module

Essay question:

Countries that intervene in trade have cited political arguments for their actions. To what extent are these arguments justifiable? Discuss with reference to examples and rationales of these arguments.

Countries that intervene in trade have cited political arguments for their actions. This essay will explore some of these arguments, and assess the extent to which these arguments are justifiable. Each argument will be presented and discussed with reference to examples and an analysis of the motivations of countries.

Despite most countries claiming to be committed to free trade (trade without intervention), they need to intervene in order to "attain political benefits" (Daniels et al. 2009: 305). Countries cite multiple reasons for intervention; a common one being that intervention will protect jobs and industries (Hill 2014).

The threat, from a developed nation's perspective, would be that jobs and industries would be threatened by low-wage (developing) countries. An example of this is the USA attempting to protect its steel industry, and its workers in 2002. President Bush introduced a tariff (a tax on imported goods) on steel imports; encouraging consumers to buy from US producers.

It could be argued that the motivations of the country were justified, as ultimately jobs were created at US steel producers and their sales revenues increased. However, the price of steel rose by up to 50% (Hill 2014) so consumers ultimately lost out. Losses to steel consumers outweighed the gains to steel producers and this highlights how some types of intervention, such as the introduction of tariffs, can reduce efficiency; and when intervention benefits one party and causes loss to another, although intentions may have been good; perhaps it is not entirely justifiable.

An example of an industry which some may argue would benefit from intervention is the UK car industry. With production moving increasingly to developing nations (Hill 2014), where labour, and thus, production costs are cheaper; the UK should perhaps consider protecting the industry. It could do this by introducing local content requirements on producers. For example, a requirement that 40% of a car's parts be produced in the UK, if it is to be sold in the UK. However, it must be considered that despite the benefits of protecting domestic job and safeguarding the industry from foreign competition; consumers would ultimately have to pay higher price for the car (to recover

higher production costs in the UK). It can be said that this is a potential drawback to intervention and leads to the question of whether intervention is really defensible.

In the end, as Hill (2014) suggests, the motives of politicians must be considered. High unemployment and instability are usually threats to the government of a country. So perhaps when they intervene, efficiency is not at the heart of their motivations, perhaps they are merely trying to ensure their political longevity (Daniels et al. 2009).

Another political argument that countries cite for their actions when intervening in trade is in order support their foreign policy objectives. "A government may grant preferential trade terms to a country with which it wants to build strong relations", or to punish/isolate countries with which relations are poor (Hill 2014: 223). An example of this is relations between the US and Cuba. The US imposes sanctions against Cuba in the hope that it will lead to economic hardship, and ultimately the fall of the government; to be replaced by a democratic pro-US one (Financial Times 2010). Similarly, it imposes sanctions on Iran in the hope that this will dent its production of nuclear weapons. So, the US is intervening in trade with the motive to influence other countries to behave in a manner that favours them. In the case of Cuba, sanctions appear to have had an effect; with improvements being made to how the economy is run since 2008 (Wild 2012), so it can be said that US intervention is justified.

Rodrick (2001) highlights how competitiveness in the world market is so important to countries, and anything that could counter this will gain the full attention of policy makers. In the light of this, it could be said that in the ongoing crisis in Syria, the sanctions placed on exports of dual use items and chemicals to Syria by the European Union since 2011 (Gov.uk 2014) could eventually have such a severe impact that its government will be forced to cooperate with other countries in order to survive. This illustrates how intervention can help to achieve political objectives, and when it makes an impact, it can be described as being one of the more valid arguments for intervention.

Countries may also retaliate to unfair competition and thus cite retaliation in order to force other countries to "play by the rules" (Hill 2014: 222) as a reason for their actions. An example of this was when imports from China (of replica products produced by US firms such as Microsoft) were breaking US copyright and patent laws. The US threatened trade sanctions against China, if China did not enforce the US' intellectual property laws (Hill 2014). As a result, China was forced to tighten enforcement of these laws, to avoid losing revenues from exporting to the US.

Threats such as this one made by the US may not always be successful, though; and it is possible that the other country may also retaliate. This would cause further issues (e.g. China could have

threatened to impose trade barriers of its own) and both countries would suffer losses. However, when successful, it could be argued that argument for intervention is ultimately valid; as it aims to protect domestic firms and prevent lost sales revenue. Additionally, it could be said that intervention in the shape of retaliation to countries that do not trade fairly can help to create a better, more fair and competitive global economy.

Countries also intervene in trade and suggest that they are doing so in order to protect consumers. As with the horse meat scandal of 2013 (The Guardian 2014), where products labelled as containing 'beef' were actually found to contain horse meat; the French government threatened to impose sanctions-labelling the behaviour as "unacceptable" (China Daily USA 2013).

Another example is of the Japanese government banning imports of American beef after a case of Mad Cow disease was discovered (Hill 2014). It could be said that this action is justified because it protects consumers from potential harm. However, a repeat case (of the disease) was not found. So it could be argued that this intervention was more harmful than good, as beef prices rose in Japan, hurting the pockets of the consumer. From this, it can be suggested that countries should ensure that they have a strong basis for intervention; and they should review their stance over time to avoid long-term implications.

Protecting human rights in other countries is an important element of foreign policy for many countries (Hill 2014); hence why some countries cite the protection of human rights as a basis for intervening in trade. Currently, the United Kingdom accuses Syria of human rights violations, and is therefore imposing sanctions on trade to the country. On the one hand, it can be said that these actions are justifiable as they send out a clear message, in a global economy which relies on free trade, that countries that violate international standards and laws will be isolated and lose out. However, on the other hand, some may argue that by trading with such countries and supporting them; a trade partner can have more influence and help to improve the situation (Hill 2014). The rationale being that by helping them economically, the government may be able to improve the situation in the country.

In the end, intervening for the purposes of protecting human rights could be described as being the right thing to do in order to encourage the improvements of the standard of living throughout the world; but countries should consider whether restricting trade would be counter-productive.

To conclude, political arguments that countries cite for their actions when intervening in trade have been explored and include reasons such as to further goals of foreign policy, to protect consumers and to protect human rights. In some instances, these arguments are not completely justified. For

example, in the case of the American Steel industry, where intervention led to a rise in prices for consumers; and when intervention leads to inefficiency. However, some reasons of countries are wholly justifiable. For example intervening in trade with a country where there is conflict present (or a country which poses a threat of conflict) is reasonable in order for a country to protect itself. Ultimately, it is important to note that some governments may intervene for the purposes of ensuring their political longevity and the best outcomes are not always gained as a result of this.

References

Daniels John D. Radebaugh Lee H. Sullivan David P. (2009) *International Business: Environments and Operations*, 12th edition, London: Prentice Hall.

Financial Times (2005) 'US-Cuba relations face setback after midterms', 10/11/2010, available at: http://www.ft.com/cms/s/0/01c2ba8a-ebf7-11df-b50f-00144feab49a.html#axzz3JWgdgJvt [accessed 23/11/14].

Gov.uk (2014). Embargoes and sanctions on Syria,2014. Available at: https://www.gov.uk/sanctions-on-syria [accessed 20/11/2014]

The Guardian (2014) 'Horsemeat scandal report calls for urgent and comprehensive reforms', 04/09/14, available at: http://www.theguardian.com/uk-news/2014/sep/04/horsemeat-scandal-report-urgent-comprehensive-reforms [accessed 19/11/14].

Hill, Charles W.L. (2014) *International Business: Computing in the Global Marketplace, 10th edition*, Washington: McGraw-Hill Irwin.

Rodrick, Dani (2001). Trading in illusions. *Foreign Policy*, March-April, P. 52-62.

Wild John J. and Wild Kenneth L. (2012) International Business: The Challenges of Globalisation, 14th edition, London: Pearson.

(source: University of Bradford (2016). Sample Essay-UG Global Business Environment Module, with ELS comments. Available at https://www.brad.ac.uk/management/media/management/els/Global-Business-Environment-sampler-essay.pdf)

Appendix B Marking Criteria

Mark & Descriptor	Key Characteristics of the Work
Distinction **Work in this category will demonstrate some or all of the following:**	• exceptional knowledge and understanding • excellent use of language and vocabulary • a very high level of critical awareness • selection of credible and relevant material from sources • excellent structure and clearly developed argument • highly effective approach appropriate to the given task • clarity, cohesion and concision
High Merit **Work in this category will demonstrate some or all of the following:**	• very good knowledge and understanding • very good use of language and vocabulary • a high level of critical awareness • selection of credible and relevant material from sources • very clear structure and consistent argument • effective approach appropriate to the given task • reasonable clarity, cohesion and concision
Merit **Work in this category will demonstrate some or all of the following:**	• good knowledge and understanding • good use of language and vocabulary • reasonable critical awareness • a good attempt to select material from sources • clear structure and argument. • suitable approach to the given task • reasonable clarity and cohesion
Pass **Work in this category will demonstrate one or more of the following:**	• limited knowledge and understanding • inaccuracies in use of language and vocabulary • basic critical awareness • some attempt to select material from sources • weak structure or inconsistent argument • failure to fully address the given task • lack of clarity and cohesion
Fail **Work in this category will demonstrate:**	• very limited knowledge and understanding • poor use of language and vocabulary • lack of critical awareness • failure to select material from sources • lacks structure and logical argument • failure to address the given task

Appendix C Semi-structured Interview Protocol

Background questions

Where are you from? What's your first language?

How long have you been working as an EAP teacher in your university?

What course do you teach in your university?

Questions about the essay sample

1. What are your opinions of the essay sample?
2. What do you think about the structure of the essay?
3. What kind of student image do you have in mind when you are marking this work?

Questions about general marking practices

1. When you are going to mark a pile of students' essays, how do you decide the marking order?

2. How do you manage your marking load?

3. When you are marking a student's work, what reading steps do you take?

4. How do you assign a score to a piece of work? Do you find assigning a score difficult?

5. How do you determine the level of critical thinking?

6. How are students who get distinction different from students who get merit?

7. What are the key characteristics of the essay for you judging its quality?

8. What do you do with grammar mistakes while you are marking?

9. What do you do if you find it hard to understand certain sentences? (for example, guessing its meaning?)

Appendix D Interview transcript of teacher A

Notes:

1.To protect the participant's privacy, answers to questions about her personal information are eliminated.

2. Lily (L) stands for me, teacher A (A) stands for the participant.

Lily (L): What's your opinion of the essay that I sent to you?

Teacher A (A) : My overall impression is that I gave the student a merit on the assessment because I feel the summative statements on the merit category fit into the essay quite well.

Overall, I think it was coherent, but structurally it doesn't meet my expectations in terms of paragraphing. I would make the statement that the cohesiveness of the writing was flawed. That meant I struggled to really get into the argument to start with because I was distracted by the structural flaws. However, once I got grips with that, it justified my expectation and some of the later paragraphs were very effective in terms of building the argument.

Overall, structuring it was lacking in cohesiveness, but internally some of the paragraphs were pretty good in terms of cohesion and developing links between ideas. There was some counterargument. That was good in building critical development of the argument so the student was being critically reflective. I also saw some evaluation in terms of final evaluation in body paragraphs. That met my expectations for an academic essay.

Structurally, I think the introduction was very poor. I didn't meet my expectation at all as an intro. It took me a while into the essay because I didn't have sufficient guidance in the intro. However, the conclusion was very helpful because it helped me fill in some gaps that I had remaining from the beginning.

It seemed to me that the student developed proficiency in their writing overall as they continued through the essay, but at the beginning it was flawed. I thought there really should be definitions paragraph because there were definitions that were being introduced on sentence level. If there had been definitions paragraph, they should provide clarity, concision, not just to initiate the reader which is me as a non-subject specialist. But at the same time, it would've allowed causal connections, links of ideas just to be much more clear and explicit.

In terms of the language, I thought it was very good, excellent. There were a couple of punctuation errors, nothing major. There were a couple of incidences of informality. The language is suddenly informal but only very occasionally.

There was lacking concision. There were lacking some more normalised phrases. Some more concise writing would've just developed the level of proficiency in language and grammar. Sometimes when I was reading a descriptive section, the student was being overly explanatory. I wanted to move on a bit quicker so that also impacted on the cohesiveness of the writing. Not only structurally but also linguistically it could've been better with cohesion.

As for citations. There was ok. There was definitely a predominance to have non-integral citations. That's fine. But it did forge the effect that student was not fronting or foregrounding the support. You had to dig to find the support. I guess my expectation as an EAP tutor of the writing structure would be that I'd advise the student to be more explicit of foregrounding the prominence of the sources. But I think the citation seems to be more accurate. They used the page numbers for

quotations so I didn't see the citations were flawed but the tone of the writing would have been improved if they've used more integral citations.

The references were consistent but I don't know the reference system the management was using. It's not Harvard, not APA.

L: What do you expect to see from somebody's intro?

A: The intro, whether in writing or presenting, is extremely important. What happens is that if I don't get guidance I need from the intro, it will ultimately affect the student's mark. This is extremely important to get it right.

First of all, I'm expecting to see the general statement which introduces the topic, using some key words from the question. The student here has only repeated one line from the question so that general statement is not satisfactory for me. They have a general statement and then they have thesis statement. But the thesis statement isn't effective either. The student is telling me how they will do but not what.

I'm looking for general statement, contextual definitions, rationale, overview of the argument, problems to be discussed, and then a strong thesis statement that shows the writer's position.

L: But the conclusion makes you clear?

A: Yes, if only the student had flipped the conclusion to the top, that would've been effective. The conclusion helped me fill the gaps because the conclusion is much more of a standard in terms of repeating the position, giving an overview of the arguments, and talking about the future.

L: The paragraphing is not good, but the essay is still coherent?

A: Here is the thing: coherent is something that I as a reader gain from reading. Cohesive is the writer has to do, to connect ideas in order to give the impression to me that it is coherent. If we think of the coherence being logic, clearly the argument is logical, and it's all about the question. They haven't gone off the topic. When I say coherence I mean logic to the argument. It starts at the beginning and he/she follows through the justifications and gives the examples of rationale. They're clearly providing a logical answer. But the problem is with the EAP tutors of marking this is we're trained to try very hard to find a logic in the argument. We're in some way biased as to assessing the logic and coherence of writing because that's our job. If you give this essay to somebody else who doesn't have that training, an academic tutor to ask them if it is cohesive, it might be a different answer. it might be the same answer. But I'm aware of the fact that I try hard to find the logic.

There're paragraphs that are much more cohesive than others, but there are paragraphs, especially at the beginning, paragraph 1,2,3. I couldn't grab the hold of what they're talking about at all. Therefore, they're undeveloped and not conventionally structured. It's difficult for me to work out what's happening. But my overall impression which you ask me at the end that the argument is logical, and there's a coherence to the argument. That improves structurally. But the cohesiveness of writing is impacted by the flawed intro and 1st couple of paragraphs.

L: What kind of student image do you have in mind when you're marking?

A: I'd say a first-year according to the confidence they appear to have in their writing. Not a brand-new 1st-year student, but not a 2nd-year student. I feel they were approaching the topic broadly. They were referring to the examples but they were quite generic examples, possibly examples you and I would have access to if you and I did background research.

L: When you're marking a class of student essays, which one do you start with?

A: If I know students through its course, then I'll choose the student whom I think has done a good job. I'll start with theirs first. Then I'll go to a student whom I think possibly is the lower end of performance and I'll do theirs. If that works out as expected, if that meets my expectation of both high and low achievers, I'll just mark as they appear on the page. When it comes to pre-sessional students, I'll just start at the beginning and work through. If I have a large class and I have less developed relationship with students, I'll just start at the beginning and work through.

L: Do you have time pressure for your marking work?

A: As EAP tutors, because our motives are spit between giving feedback in improving writing and then providing a grade, this is more challenging and time-consuming. Over the years of marking, I've got quicker and I've developed strategies in order to work smarter to keep to that time frame. That's a conflict of me as an EAP tutor whereby you want to give lots of detailed information but in actual fact it's detrimental to your own working life and it's also to be detrimental to your students' ability to take that on board and action it. I think targeted feedback is always better than comprehensive feedback.

Unless it's clear from the week before we're focusing on what is writing, I'll always say at the top of my essay marking "today, I'm focusing on organisation and argument." "today, I'm looking at language and grammar" or I'll section a piece of essay, look at language and grammar, and section a piece of essay to look for cohesion, or citation. So students know what approach I've taken.

L: When you're marking a student essay, what reading strategies do you take?

A: I think I've developed strategies for saving time. I look at the structure first of all, so I look at the macro organisation and structure, and I look at the cohesion of argument, and then finally I look at some language features. Language features and grammar are my final consideration. I'll look at the structure, I look at the intro and conclusion, and I pick a paragraph and look at how they developed the argument. Then I have an overall sense. I'll go through for citations to see how they've done that, and then skimming I pick up language features as well.

L: Do you find it difficult to give a score to a student work?

A: It's always trying to match the criteria to my thoughts about the essay which takes time because I have reading strategies. It always takes time to try to decide which of the criteria descriptors is going to fit my thought of the essay. Of course, ultimately as an experienced marker, you have just to get a sense that it is a B or D. Mostly, it is fine but in some cases where it could be C or D, it could be a B or A. It's really tricky. This idea of whether it's an A or B by just flickering through it. That's making assumptions about students which I feel unfounded. I feel each student requires the same amount of care and attention of their writing.

Even when I was in a university, tutors would say the intro is very poor, the 1st page is very poor because I'm making assumptions about your overall grade based on that 1st page and I used to think "will you? That's not very fair, isn't it? That student is a case in point because they start off very poorly but in my opinion, it improves as it got on.

I try not to do that, because I think that's an unfair approach, but I'm sure I probably do subconsciously.

L: What do you do with student's grammatical mistakes while you are looking at their essays? Do you correct all of them?

A: It depends on the nature of mistakes. Something like that informality of "so", I think that's important from that stylistic point of view or the student writes "in the end", that's way too informal. When the formality is style, I feel very quick to pick up on. If the grammar mistakes are systemic, so they're running all the way through, I won't make a comment on each and every one. I'll make a comment in the end and say "your article use is rather distracting. work on articles." But I was not going to pick out those errors as we go. If it's not systemic, I'm not going to make a comment on it. I won't comment line by line if the meaning-making breaks down. If the communication breaks down, then I write "I can't understand your point due to your grammar."

L: Do you try to guess what your student is trying to say when you don't understand?

A: No, because I don't think that's help for the student. They SPS tutor won't try to guess. Their SPS tutor would assess based on what they can understand. Therefore, I don't think it's helping our students coming for support or going to go on writing in the university.

L: How do you understand critical thinking?

A: Critical awareness is in 2 parts: there' breadth of reading, so there's how well all ideas are supported on the topic, how much the student has read. Then there's critical awareness in terms of writing. Are they doing critical writing? Are they providing a claim? Are they explaining the claim, giving an example, then giving a counter-point to that, then refuting that? Are they giving transitional sentences to evaluate? Do they have a final comment? Having broad understanding of different theories and implications, different school of thought, and then applying it to writing, showing that they can develop an argument.

I think they've done that in the first instances of the claim better than providing a counter-point so give a counter-point but that's what they've done. They haven't refuted it. To some extent they've done critical writing because they have a claim and counter-claim, but they haven't got a refutation. I would expect to see that interplay's going on very well-developed piece of writing.

L: What are the most important characteristics of the essay for you to judge its quality?

A: The key features are structure and organisation, developing the argument, cohesion and coherence, language use, referencing and citations.

L: How are students who get A different from students who get B?

J: Generally, an A-grade student, a high-achiever is doing things perhaps not structurally better but stylistically better. Structurally A and B essays could be the same, but the style is more developed. There will be a strong thesis statement that has a position and through that guides me completely through the writing. There would be using language features like normalisation.

Back to these three things: clarity, concision, and precision. There would be showing clarity in their language, position through writing, and concision. You could find these 3 themes in structural elements but that's not going to make such difference between A and B, make a difference between B and C.

At A and B level, these three themes are much more proficiently done.

Also, the reading would be effortless. It would require much effort to be able to clearly get the points.

L: Do you give any tips to your students how to develop a clear position?

A: We spend a lot of time on thesis statements. If from the beginning, the reader, the marker is not clear about your position, doesn't know not only where you are going but also how you're going to get there, and doesn't have that knowledge to begin with, then you can have really well-developed arguments but don't fall under anything. You can have arguments and you can see perhaps the lateral connection between the arguments in the body, but there's no vertical connection between those arguments and the central point that's being made in the thesis. These things are coherent and cohesive, but they don't connect up to the central point. For an A level, they must. We spend a lot of time on thesis statements.