University of Southampton Research Repository

University of Southampton

Faculty of ARTS AND HUMANITIES

MODERN LANGUAGES AND LINGUISTICS

Capturing the Human Dimension of Performance Assessment: An Investigation of Decision Making in Academic Oral Presentation Assessment Tasks on English for Academic Purposes Courses at UK Universities

by

Louise Claire Palmour

ORCID ID: 0000-0002-3426-8414

Thesis for the degree of Doctor of Philosophy

April 2020

Abstract

With renewed attention on graduate attributes and active learning in UK Higher Education (HE), academic oral presentation (AOP) tasks have become popular learning and assessment tools on degree programmes at UK universities. On English for Academic Purposes (EAP) programmes, AOPs often function as gateway assessments, which may grant or deny students access to university study. This thesis reports an investigation of decision making in such high-stakes assessment tasks on EAP programmes.

Due to a paucity of research on AOP assessment practice in HE, I conducted The Survey Study. This survey yielded data, primarily through document analysis and questionnaires, on subject module and EAP module AOP assessment tasks and practice. The Fieldwork Study was then conducted, which involved accessing teacher and student AOP assessment practices in two EAP settings. To capture data on the AOP assessment processes, I used ethnographic data-gathering techniques including document analysis, observation, interviews, and field notes. Both phases of the project draw on the inquiry traditions of constructivist grounded theory (CGT) and ethnography and use CGT as an analytic framework.

The Survey Study findings paint a picture of diverse applications of AOP tasks on subject modules and EAP modules. The findings indicate that delivery, as well as content, is often key to successful AOP performances. From The Fieldwork Study data, I developed a theory which represents the complex mix of processes EAP teacher assessors enact in their practice linked to an AOP assessment task. The EAP practitioners are *Inheriting and Influencing AOP Assessment, Constructing AOP Task Representations, Communicating AOP Task Representations, Gathering and Selecting Assessment Evidence* and *Critically Reflecting on AOP Assessment Practice*. When reaching score-decisions, teacher assessors employ a zoom function, a number of lenses (Judging, Marking, Describing, Evaluating, Technical, Aesthetic), and filters (Criteria, Ranking, Social, and Interpersonal) to the performances. The findings demonstrate that the social context, which includes practitioners' histories and values, influence teachers' practice in particular.

Through an investigation of student perspectives on AOP tasks, I demonstrate that a range of factors (Personal, Content, Language, Assemblage, Engagement, Affective, Formative and Summative) shape students' thinking and action. Importantly, students devote energies to showcasing content-related skills, rather than to linguistic performance in AOP assessments. Furthermore, a number of students downplay the importance of spoken language in AOPs which seems to be due to the ability to draw on a range of semiotic resources to achieve a successful performance in their contexts.

The Survey Study and The Fieldwork Study findings provide a nuanced conception of what AOP tasks may involve and support the case for intensified research into the communicative characteristics and demands of such tasks. The processes and strategies illuminated in The Fieldwork Study critique the notion that there is an assessment format which operates independently of input from teacher assessors and student test takers. The assessment is instead realised and constructed in its implementation. The teacher assessors' and students' substantive processes in AOP assessment have ramifications for how validity is conceptualised and enhanced in AOP assessments in EAP contexts.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: LOUISE CLAIRE PALMOUR

Title of thesis: Capturing the Human Dimension of Performance Assessment: An Investigation of Decision Making in Academic Oral Presentation Assessment Tasks on English for Academic Purposes Courses at UK Universities

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: Date: 5th May 2020

Acknowledgements

I wish to recognise all those who have supported and encouraged me throughout my doctoral studies.

First, I would like to thank warmly Dr Richard Kiely, my supervisor. Richard has been unwavering in his support and kindness. Always encouraging, Richard gave me a perfect balance of invaluable guidance and freedom to have ownership of the project.

I also wish to thank my family and friends. I feel fortunate to have you by my side. I am grateful to my mum and dad who have been a constant source of encouragement. Laura P has been the most caring sister and flatmate: thanks for the gigs and jogging. To all my friends, you lift my spirits in so many ways. A special thank you to Alfie Santos and Katerina Harris.

I also would like to acknowledge my funders the Economic and Social Research Council and the South Coast Doctoral Training Partnership, particularly for the freedom to conduct the project as I wished.

Thank you to my examiners for their feedback and the friendly and stimulating discussion in the viva.

And finally, a huge thank you to those who have participated in this study, especially to the teachers and students at the two field sites. I have been heartened by your willingness to share your worlds with me; I have learned so much from you.

I dedicate this thesis to my Grandad Harry who taught me to marvel at the joy and opportunities education can bring.

List of Acronyms

AOP   Academic Oral Presentation

CBA   Classroom-Based Assessment

CGT   Constructivist Grounded Theory

EAP   English for Academic Purposes

ELF   English as a Lingua Franca

ELFA   English as a Lingua Franca in Academic Settings

EMI   English Medium Instruction

EP   EAP Practitioner

EPQ   EAP Practitioner Questionnaire

FS1   Fieldwork Site One

FS2   Fieldwork Site Two

GT   Grounded Theory

HE   Higher Education

IELTS   International English Language Testing System

IBP   Interview before Presentation

IAP   Interview after Presentation

IFP   International Foundation Programme

S1   Study One

S2   Study Two

SL   Subject Lecturer

SLQ   Subject Lecturer Questionnaire

Glossary of Terms

The purpose of this section is to act as a priming for the reading of this thesis. It functions as a glossary of important terms and concepts which can also be referred to retrospectively.

Assessment and Test

I use the words assessment and test interchangeably in the thesis. However, the word assessment features more prominently in the text to follow. This is because the term assessment is often presented in literature on classroom-based assessment (see Rea-Dickins, 2001 and Hill and McNamara, 2011) as an inclusive term suited to such fluid classroom settings and less standardized assessment decision making which feature in the settings under investigation in this project. I conceptualise assessment in line with Hill and McNamara's definition:

> any actions, interactions or artifacts (planned or unplanned, deliberate or unconscious, explicit or embedded) which have the *potential* to provide information on the qualities of a learner's (or group of learners') performance. (p.397; italics in original)

The definition is broad in order to capture the actions of teachers and students which take into account the 'affordances' (Van Lier, 2004) in learning and assessment activities. Affordances are the 'possibilities for action that yield opportunities for engagement and participation' (p.81). A rich array of assessment opportunities are available to teachers and students throughout the lifespan of the educational courses which feature in this research.

When referring to work of authors, I use the term test or assessment which they adopt in their original text.

English for Academic Purposes

Any module which cites (for example in module descriptions) as a primary explicit purpose the development of English for academic study is regarded as an English for Academic Purposes (EAP) module in this project. This includes, for example, such modules which are included in year-long International Foundation Programmes, Pre-Master's courses and summer pre-sessional courses.

EAP does not entail a prime, or isolated, focus on language development, despite the development of English being the privileged property I have used to select field sites in this study. EAP practitioners draw on diverse theory in their teaching and research activities, and their practice often centres on developing discourse competence, recognising the importance of social practices, disciplinary epistemologies and ideological beliefs in communication (Hyland 2018, p.12). EAP teaching often involves a focus on academic literacy and skills development (writing essays, preparing and delivering presentations, participating in seminars) rather than a focus on

linguistic features such as vocabulary and grammar. (Schmitt, 2015). (See Section 2.4 for a detailed discussion on the construct of EAP).

Individual actors may not conceptualise the modules featured in this project as EAP modules, or informants as EAP practitioners. However, the respondents chose to participate in a study which was described as an investigation of EAP practices. I acknowledge that using the term EAP has disadvantages as it may prompt readers to bring their own associations of this construct to their reading of this study. Part of this study's agenda is, however, to probe existing conceptualisations of the construct of EAP.

ELF

English as a Lingua Franca (ELF) communication is described as 'any use of English among speakers of different first languages for whom English is the communicative medium of choice' (Seidlhofer, 2011, p.7). Widdowson describes contemporary scholarship, which researches this phenomena, as directed at investigating variability of English in different social contexts between those who potentially do not have linguacultural commonality. He describes ELF communication as an ideal site to understand the on line processes of variation and to examine the legitimacy of variability (Widdowson, 2020, p. 172).

The third phase of ELF ('ELF 3'), which locates ELF within a framework of multilingualism, sees ELF 'defined not merely by its variability, but by its complexity and emergent nature' (Jenkins, 2015, p.77). ELF scholars differ on conceptualisations of how emergent ELF or 'English' is. Widdowson proposes that English is recognisable due to a 'virtual language', which has infinite realizations but grounded in a fixed code; whereas other ELF scholars regard realizations of ELF as emerging from interaction instead of reference to an abstract system (Baird et al., 2014; Jenkins, 2015). An emergent position, guided by complexity theory, expounded by Baird et al. (2014) particularly underscores the importance of examining ELF use in relation to practices and context.

This project retains a primary focus on capturing salient processes teachers and students enact while engaged in AOP assessment practice. ELF research is drawn upon to probe orientations towards perspectives on language use and communication in EAP assessment performances, instruction and assessment decision making. In particular, ELF scholarship has relevance to exploring understandings of what are deemed appropriate approaches to EAP assessment constructs. ELF research which informs the project is covered mainly in Section 2.9.2.devoted to critical perspectives on language assessment.

EMI

Degree courses at UK universities are conceptualised as EMI programmes in this study. I class EMI programmes as those in which an academic subject is communicated through the medium of English in a tertiary educational setting. EMI programmes are ones in which English is predominantly used as the medium, not the target, of instruction, with minimal explicit reference to language learning aims on these programmes (in line with Schmidt-Unterberger's 2018 EMI paradigm). Anglophone institutions are excluded by some scholars from EMI definitions (e.g. Macaro et al., 2018) and research. However, recent work in EMI (Baker and Hüttner, 2019; Jenkins and Mauranen, 2019) has argued that Anglophone settings be included under EMI because the Anglophone contexts share similarities with non-Anglophone contexts. UK universities are international, multilingual sites (Jenkins, 2014) with staff and students using multilingual repertoires. Baker and Hüttner (2019) convincingly argue that concerns about linguistic proficiency, language practices and their influences on learning and teaching are as significant here as in non-Anglophone settings.

Degree programmes, and subject modules (see term below in glossary) within these, which feature in this study are conceptualised as EMI programmes. Pre-sessional EAP programmes in this research take place in UK universities which are classed as English medium institutions. Pre-sessional EAP programmes (see term below in glossary) prepare students for their target disciplinary English-medium degree programmes, but have a clear separation from these.

Language+ assessment

As the project progressed, problematizing the label 'language assessment' while continuing to reference important work in language assessment scholarship became expedient. Therefore, I use the term language+ assessment, in select areas of the thesis, to signify that practices subsumed under the label 'language assessment' involve significantly more than an assessment of language. In fact, the language criterion is increasingly diluted in many 'language assessments', notably in EAP contexts.

In EAP settings, the language learning may be incidental (see more on EAP in glossary of the term above and in Section 2.4). What is gauged through instruction and assessment on pre-university EAP courses is how students are able to learn through the medium of English and cope with UK HE delivery methods and assignment requirements (Schmitt, 2015). An essential point to consider throughout the thesis is how well 'language assessment' represents EAP assessment. I discuss in what ways the project findings develop conceptualisations of the EAP assessment construct and relevance of current labelling of such assessments in Chapter 8.

Performance

The concept performance has been instructive in describing, interpreting and theorising the nature of AOP communication and AOP assessment decision making. The term performance is used to highlight the social quality of the actions and events in this study.

It features prominently because the AOP assessment tasks are considered to be second language performance assessments. 'Performance' entails that 'actual performances of relevant tasks are required of candidates, rather than the more abstract demonstration of knowledge' (McNamara, 1996, p.6). On occasions, I therefore refer to the loosely framed social events of the AOP assessments as 'the AOP performances'. I discuss definitions, characteristics and key issues in performance assessment in Section 2.8. Furthermore, I consider communicative acts as performative in nature meaning that I align myself with the emergent perspective which maintains that competence does not necessarily prefigure performance. Conceptualisations of the relationship between performance and competence in communication are explored further in Section 2.8.3.

Pre-sessional course

I use the term pre-sessional course to refer to EAP courses (see glossary of term above) which take place *before* students enrol on their disciplinary degree programmes and have a clear separation from the target degree programme. I include year-long programmes such as Pre-Master's and International Foundation Programmes in this category as well as shorter summer courses. I acknowledge that the courses grouped in this category may be grouped differently in the institutions in which they are managed and that the practices on courses grouped under this label will operate in diverse and often very distinct ways.

Student test takers

I do not refer to the students in this thesis as test takers as their identities as learners are central to their assessment decision making. I, instead, use the term student test taker.

Subject module

I use the term subject module to refer to taught undergraduate and postgraduate modules on degree programmes at UK universities. I also use the term disciplinary modules in parts of this

thesis as a synonym. Subject modules are distinct from Pre-sessional EAP modules (as discussed in the section on EMI above).

Subject lecturer

I recognise that many staff at UK universities who assess oral presentations on taught undergraduate and postgraduate courses will have different job titles to that of lecturer. I include all teaching staff members in this description, and hope that it is not in any way perceived as excluding particular teaching staff. I also use disciplinary tutors to refer to teaching staff on subject modules. HE practitioner is used to refer to both subject lecturers and EAP practitioners.

Teacher assessor

I refer to the HE practitioners (EAP practitioners and subject lecturers) who feature in this study as teacher assessors due to the interwovenness of their teaching and assessment practice. I am primarily investigating teachers' role in assessing students, however my time in the two fieldwork settings shows that their actions as assessors are inseparable from their role as teachers. Throughout this thesis, I choose not to refer to the HE practitioners as assessors without acknowledging their pedagogic identity.

Stylistic Choices

Transcription conventions

.	Leading to end of utterance
,	Slightly rising tone giving a sense of continuation
" "	Reporting speech or text (e.g. from documents, peers or teachers)
?	Rising intonation
!	Animated tone
CAPS	Louder than surrounding talk, indicating emphasis
( )	Contextual information including features such as body language and notable changes in informants' tone
[ ]	Overlap
{ }	Clarification by researcher
{...}	Parts of transcript removed

I conducted predominantly content analysis on the spoken data, therefore I have not included a range of features which may be required in a conversational analysis. I included such features as tone, overlap and capitalized for emphasis so as not to erase the emotive intention in the speech. This information has served to strengthen the analysis, namely the plausibility of interpretations. I transcribed the spoken data into intelligent verbatim, which involved removing some reformulations, false starts and repetition. It was not deemed necessary to alter any non-standard language use, however, as the utterances proved sufficiently intelligible. Furthermore, to take such action would erase the unique and diverse linguistic repertoires and modes of expression of the informants. I also believe that the promised level of anonymity is not compromised in retaining this linguistic diversity.

Data identifiers

For Study One data, the source is acknowledged after the data extract. I include the respondent number (e.g. SL5 or EP5) at the end of data extracts from Study One questionnaires. For Study Two data, I reference the field site, method and informant pseudonym (e.g. FS1, IBP with Aimee). The acronyms used are listed above in the list of acronyms.

Questionnaire data presentation

I have presented questionnaire data as it was written by respondents. The hierarchy has been preserved and any typos have not been corrected.

Chapter 1   **Introduction**

1.1   **The Context**

> At the back of your mind is you're the gatekeeper to their undergraduate programme. If you're passing them, are you sure that they are going to be able to perform on an undergraduate programme. You know I'm going to let these students loose. Are they going to be able to cope?
>
> (Tracey, teacher assessor on an International Foundation Programme)

Tracey shares what ultimately drives her decision making when assessing summative presentation assessment tasks on an International Foundation Programme (henceforth IFP) at a UK university. The IFP is a year-long programme for students who wish to meet entry requirements to enrol on an undergraduate degree programme. The academic oral presentation (hereafter AOP) assessment task is weighted at 25% in a suite of assessments on the IFP which determine whether students commence their prospective disciplinary studies in the following academic year. Tracey, along with a co-assessor, makes a judgement as to whether a student she has taught for an academic year will 'be able to cope' on an undergraduate programme from an AOP task among other assessment opportunities.

Currently, many UK universities make a distinction between 'home' and 'international' students in their recruitment and admission policies. In order to gain entry onto a course of study at a UK university, many prospective 'international' students are required to meet English language entry requirements. A student may be accepted on a UK degree programme by sitting, and obtaining the required result in, a large-scale English language proficiency standardized test, such as IELTS. Alternatively, a student may, in many cases after not gaining the required result in large-scale tests, enrol on a pre-sessional course and complete in-house university assessments. While large-scale standardized tests used for university admission purposes centre on assessing English proficiency, others have a stronger performance element which involves the deployment of an array of non-linguistic skills on the part of the test taker (Hamilton et al., 1993). The number of students taking such in-house pre-sessional gateway assessments in UK institutions is not clearly documented. In the institutions researched in this study, the numbers of students taking such assessments on pre-sessional course in one institution was a maximum of 20; in the other institution the number has been over 1,500 students per year for the past few years. Each year across the UK thousands of students sit in-house pre-sessional assessments.

Devolved, in-house pre-sessional instruction and assessment offer to equip students with the 'communicative readiness' to embark on their target disciplinary studies through the medium of

English. Pre-sessional courses are targeted at students wishing to be accepted onto a degree programme, as opposed to in-sessional courses which support students currently completing their disciplinary studies. These pre-sessional assessments take place after a course of instruction making these classroom-based educational assessments. Educational assessment has the opportunity to capture a more holistic interpretation of a student's overall potential to perform in a range of dimensions (e.g. linguistic, cognitive, social and emotional) in their host academic communities. What teacher assessors and student test takers do, experience and value in devolved assessment, in AOP tasks in particular, has been little explored. This is surprising given the wealth of AOP practice in Higher Education (henceforth HE), in both pre-sessional English for Academic Purposes (hereafter EAP) programmes and disciplinary programmes.

1.2   Personal Interest

During my time as an EAP practitioner, I have become particularly interested in what shapes the delivery and reception of oral performances communicated through English in academic settings. Upon embarking on this project, I wished to explore EAP practitioners' and subject lecturers' perspectives on what factors shape performances in specific genres in assessment contexts. Academic oral presentations are a key genre in academia and feature prominently on EAP courses. I was eager to begin to unravel the workings of assessed AOPs in UK Higher Education. As a practitioner, I found that marking AOPs is fast-paced and I was concerned at times at the few opportunities available to reflect on the assessment decision making. I have long been intrigued by what factors shape what I and my colleagues deem to be an 'appropriate' or 'successful' AOP performance. I am particularly drawn to how EAP practitioners' treatment of language and communication interacts with other aspects assessed and how compatible it is deemed to be with individual AOP task characteristics. These areas of particular curiosity prompted me to find out more about the different purposes and functions of AOPs used on degree programmes and EAP programmes. I wondered how teachers and students set, navigate and challenge the requirements and expectations associated with the act of giving and marking a presentation. Developing a closer understanding of AOP practices seemed especially important in the EAP contexts in which high-stakes decisions are taken as to whether students are granted access to study at a UK university, based partly on AOP performances. Assessed AOPs became events of interest because the assessment processes are potential windows on what truly counts to students and teachers. I wished to gain insights into practices and first turned to literature, but a resultant scoping review proved that work conducted into EAP AOP assessment is lacking.

1.3   The Research Problem

EAP programmes aim to prepare students for practices on their target degree courses in English medium instruction (hereafter EMI) institutions. EAP curriculum and assessment are presented as flowing from needs analysis, and rights analysis (Benesch, 2001), informed by practices in host academic communities. In UK universities, oral presentation assessment has become increasingly common (Joughin, 2007). Therefore, oral presentation assessments are often used as EAP assessment tasks on pre-sessional courses (see Chapter 4 for evidence to support this from this project). Despite the wealth of AOP practice in HE, little research has been conducted on AOP tasks used as assessments on taught courses, particularly the communicative demands involved in such tasks (Section 2.3 includes a summary of studies conducted).

AOP tasks on EAP courses which act as gateway assessments have the power to determine international students' entrance to university. Such assessments have huge social consequences (Spolsky, 1997), but there is a dearth of research which has theorised the decision making processes in such gateway assessments (see Chapter 2 for discussion of specific lacunae). Without such a knowledge base, there is a risk to the quality of assessments. There is a danger that what is happening in these assessments is not sufficiently visible and therefore being under-represented and perhaps even misrepresented.

In this research, I examine AOP practices on subject modules and on EAP modules. Pre-sessional courses aim to act as bridges to practices in host academic communities. Therefore, I gathered information on AOP tasks on degree programmes with the primary aim of facilitating the understanding of decision making in EAP practice related to AOP assessment events.

1.4   The Study

The project comprises two studies which obtain and analyse predominantly qualitative data at universities in the UK: Study one (henceforth S1) and Study two (hereafter S2). Core details of the research are provided in Table 1 below.

Study name	Study one (S1) The Survey Study	Study two (S2) The Fieldwork Study
Description	Qualitative surveys documenting AOP formats and stated practices on subject programmes and EAP programmes at UK universities.	Qualitative fieldwork investigating AOP assessment practices on EAP modules at two sites.
Purpose	S1 gathered information on factors which shape AOP purposes and communicative demands in UK HE. This is used to inform the focus of fieldwork in study two.	S2 obtained and analysed data on EAP AOP practices. In doing so, the research provides accounts of teacher and student decision making in AOP assessment tasks.
RQS	1. In what ways are AOP assessments used in UK Higher Education? 2. What is the EAP AOP assessment construct? 3. What factors determine the EAP AOP assessment construct?	

Table 1: Research Overview

In this study, I blend methodologies to investigate the underexplored and complex area of AOP assessment in Anglophone EMI settings, partly on subject modules, but predominantly on EAP modules. The research is informed by principles of ethnography (Hammersley and Atkinson, 2007; 2019) and constructivist grounded theory (henceforth CGT) (Charmaz, 2014). CGT is used as an analytic framework (see Chapter 3 for a full account of the methodological approach). In the study, I have collected data using a range of sources. While this thesis predominantly reports on the data from the central data sets due to scope and space limitations (Study one: questionnaire data; Study two: observation, interviews and rating discussions), the data from all sources has informed my perspectives on the research topic. Figures 1 and 2 below represent the range of methods I used throughout the course of the research. I present these as a fan of methods (similar to Salter-Dvorak, 2010, p.26). These figures represent the effort made to engage with the topic on a range of fronts.
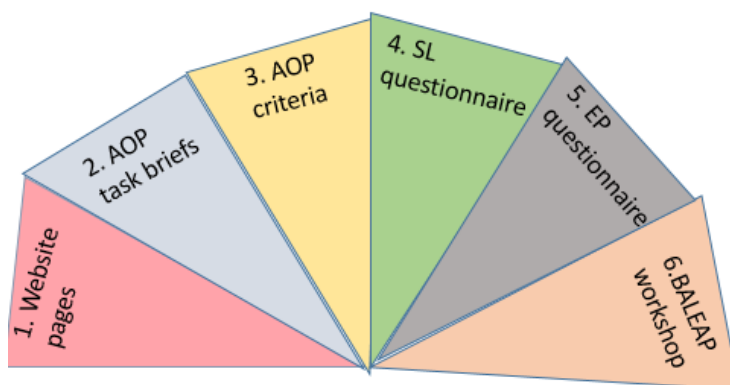
Figure 1: Study One Methods

In S1, first, I surveyed a university's web pages to develop an indication of types of AOP tasks currently communicated as in use. Second, I sourced a collection of task briefs and marking criteria from EAP and subject modules at UK universities. Third, I devised questionnaires for subject lecturers and EAP practitioners. Fourth, I conducted a workshop presenting preliminary findings from S1 at a BALEAP (an international professional network for EAP practitioners) Practitioner Issues Meeting to engage in dialogue about AOP assessment with EAP practitioners from different institutions. The BALEAP workshop was an opportunity to explore practitioner thinking and concerns related to AOP assessment, but it was not used to generate any data sets.
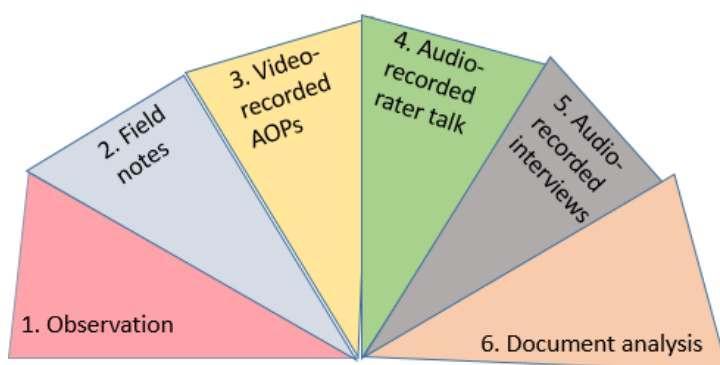
Figure 2: Study Two Methods

In S2, I observed a number of lessons and the AOP assessment events, including discussions between teacher assessors reaching score decisions. I took field notes during these encounters. Alongside this, I video recorded the AOP assessments and audio recorded rating discussions. I accessed individual student and teacher accounts in audio-recorded interviews. Document analysis of the marking criteria and syllabus was also conducted.

This study focuses on in-house, classroom-based assessments, which are often developed and administered by teachers. The teacher assessors may have greater flexibility in monitoring and assessing student test takers than in large-scale assessments (Green, 2014). In such contexts, an interpretive approach, which involves gathering data on student and teachers' perspectives and

behaviour, is well-suited to studying AOP phenomena. The triangulation of the methods (in Figures 1 and 2 above) enabled me to gain diverse perspectives and richer understandings on phenomena.

In S2, I conducted ethnographically oriented fieldwork which involved a constant negotiation of 'insider/outsider' positioning in my interactions with informants. As McNess et al. (2016, p.34) argue 'we are neither complete observers nor complete participants but often working in that 'third space' in between.' Akin to Salter-Dvorak (2016a, when describing her positionality in her 2011 study), I took on multiple identities during this research. One the one hand, I had not worked with the informants previously, meaning I could maintain a level of critical distance and may be viewed as an outsider in this respect. On the other hand, I had some experience as an EAP practitioner which created a level of insider status. I had the opportunity to share stories of similar lived experiences as a student in UK HE and EAP practitioner with the student and teacher informants. I took on emergent identities of a fellow EAP practitioner, fellow student in HE, fellow classroom participant, additional teacher, and visiting researcher (see Section 3.5 and 3.8.1 for further discussion and vignettes which describe the different roles I had in the field).

1.5   Thesis Structure

In Chapter 2, 'Theoretical Framework', I discuss a number of pivotal theories, concepts, constructs and research streams with relevance to the project. I include details on how I position this research in relation to research on oracy, EAP, language testing, ELF, performance assessment, classroom-based assessment (henceforth CBA) and validity theory. Chapter 3, 'Research Design and Methodology', includes descriptions of research procedures and the rationale behind them, as well as key details about the setting and participants.

The next four chapters in the thesis present data analysis. Chapter 4, 'A Survey of AOP Practices', is an analysis of S1 data. In Chapter 5, 'Teacher Assessor Processes in AOP Assessment', I theorise core processes involved in teacher assessor actions in the two field sites. The analysis is mainly focused on what teacher assessors do prior to and during the AOP assessment process. I discuss how teachers inherit and influence the AOP task, the ways they communicate their evolving task interpretations to students prior to and during assessment, how they gather and select evidence on students' abilities and how they critically evaluate their practice. Chapter 6, 'Teacher Decision Making in Rating AOPs', showcases the analysis of the two rating discussions between the assessors at the two sites, in which they reach scoring decisions. I present data which demonstrates that assessors employ a zoom function, various lenses (Judging, Marking, Describing, Evaluating, Technical and Aesthetic) and filters (Criteria, Ranking, Social, and

Interpersonal) in order to agree upon a mark. The final analysis chapter is Chapter 7, 'Student Perspectives on AOP Assessment'. In this part of the thesis, I present student data from the two field sites, mainly from drawing on interview data. I lay out factors (Personal, Content, Language, Assemblage, Engagement, Affective, Formative and Summative) which shape student thinking and actions related to the AOP assessments.

In Chapter 8, 'Discussion', I present the main findings, my interpretation of the findings and their contribution to established research bases. In the final chapter, 'Conclusion', I give a summary of key findings and implications of the research. I also provide reflections on the PhD process.

1.6   Summary

In this opening chapter, I have set out the research context and problem and shared my motivation behind choosing to investigate AOP assessment practice. I have also provided core details of the study's approach and methods. The next chapter discusses research streams, concepts and constructs which I key into in order to aid data interpretation and characterisation of teacher and student accounts of their experiences, perspectives and practices.

Chapter 2   Theoretical Framework

2.1   Introduction

The theoretical framework presented here has been assembled prior to, during and after the main data collection phase in this study. I use the term bricolage to describe the theoretical framework which is presented in this chapter, because the framework has been constructed from a wide variety of sources. The bricolage contained herein has been assembled over three years by means of entering into dialogues with texts, people and artefacts in a variety of settings.

In the early stages of the project, I conducted an initial review of literature. This scoping review, established the existing research base related to oral presentation assessment in HE and on EAP courses. The data then took centre stage and pointed to fruitful leads to pursue. I gave the data precedence and room to shape the study's direction by coding what was happening in the data, rather than using preordained categories or theories from Applied Linguistics and other disciplines (Chapter 3 offers a detailed description of, and rationale behind, the methodological decisions taken). I have chosen not to frame this study predominantly in one school of research, such as EAP, ELF or Language Testing. This strategy was adopted because the complex research phenomenon under investigation required the construction of a novel conceptual framework, which required concepts in a range of fields. However, the aforementioned research streams inform this PhD and it is hoped that this project has something to offer to these research areas.

I recognise that I was by no means a blank slate when conducting this research, and this had the potential to influence what theories were used to account for practices, what artefacts were included in the bricolage. My previous studies in the field of Applied Linguistics, background as a language learner, and time as an ESL teacher and EAP practitioner in the UK and EFL teacher in Russia have shaped the paths I have followed in this research. For instance, although I was interested in the role of non-linguistic factors in the project from the outset, I believe at times I treated language as the centre of gravity when exploring activities of interest, especially in S1. My previous experiences had left me predisposed towards holding what may be labelled a lingual bias, or - more generously - a leaning towards the linguistic. Further exposure to works which challenge the study of language in isolation resulted in this PhD becoming an exercise in rethinking, and to some extent dismantling, this linguistic paradigm in approaching communicative practices. When analysing the data and making judgements on people, however, it has been my belief in the power of 'quiet competence' (Cain, 2012) and aversion to practices which promote 'the extrovert ideal' (Cain, 2012) which I found could warp my reading of events. I

exercised reflexivity in order to remain open and not too readily deem certain behaviour as an affront to, or acceptance of, quieter students' behaviour in particular contexts. Recognising and exercising control over the impact of dispositional struggles in my life, as well as academic training, materialised as particularly crucial in reaching balanced interpretations of events.

I have endeavoured to ensure my rendering of events and data remained close to the essence of what was happening. One strategy to achieve this was to stagger the admittance of existing literature by interrogating its instructiveness and fit with the data. Instead of theories filtering the data at the outset, data filtered the theories. Constructivist grounded theorist Kathy Charmaz (2014) advocates taking such an approach:

> you follow the leads gained from your view of the data, not from the careful and exhaustive literature review of the traditional research design (p.27)

The literature deemed instructive for examining the research phenomena has evolved throughout the research process. This project primarily investigates EAP AOP practices on courses which are targeted at second language users of English. Therefore, this study retains a degree of focus on the role of language and communication in AOP performances. Engaging with EAP performance assessment decision making, which includes the linguistic reality in AOP performances, requires a conceptual framework which has the ability to make sense of thinking, actions and practices constituted by a complex web of interrelated factors and discourses. Baird et al. (2014) outline the need to acknowledge the complex nature of communicative performances:

> 'We need to engage with the linguistic reality in which signs, both spoken and written, integrate with numerous processes, systems, and artefacts in communicative performance, as well as the ways in which communicative performances engender aspects of simultaneous but dynamic identification, social positioning, knowledge, expectations, experience, and habit' (Baird et al., p.175)

I take a broad and holistic approach to the study of AOP communicative performances which engages particularly with integration of processes and artefacts and explores such aspects as social positioning, knowledge, expectations, and experience. Teacher assessor and student test taker informants in this study engaged in AOP tasks are concerned with communicative performance but also pedagogic, learning and assessment processes among others. In the following sections, I discuss the research bases and concepts in a range of areas all connected by their relevance to understanding teacher and student decision making in complex assessed AOP communicative performances on educational courses. The inclusion of two strands in the PhD (S1 and S2), which encompass exploring not only AOP assessment practices on EAP programmes but also on taught degree programmes at UK universities, has meant the project has relevance to many established discourses, which I outline next.

In delimiting the scope of this research, I described the research as 'a study exploring oral presentation assessment practices in UK HE'. AOPs have been widely categorised as an 'oral' genre and this proved to be the case for the informants in S1 and S2. The concept of oracy became particularly relevant. However, the notion of oracy required broadening to encompass the rich repertoire of semiotic resources (linguistic and non-linguistic features of communication) drawn upon in AOP communication in the settings which feature in this research. I first discuss the concept of oracy (in Section 2.2 below). I, then, review studies which add to a growing base of studies on oracy demands in HE and highlight the limitations in some current conceptualisations of oracy. I summarize the scholarship which has been conducted particularly related to AOP tasks on subject courses and EAP courses (see Section 2.3 below).

My research primarily investigates, and contributes to, the field of EAP, therefore in Section 2.4, I provide my inclusive interpretation of EAP. Following this, as I centre the investigation on AOPs, I discuss how this study is informed by genre analysis (Section 2.5). I devote Section 2.6 to the concept of genre and previous work which has offered descriptions of AOPs. As this study is focusing on AOPs which are used for assessment purposes, I review discourses in language testing and assessment shown to be relevant in the data (in Section 2.7). In the early stages of the PhD, I came to conceptualise AOP assessments as performance assessments, therefore I include a section (2.8) on important issues which derive from examining performance assessment. These issues which have resonance with the data in this project are: the role of the criterion, the role of linguistic and non-linguistic factors and the concept of performance (covered in Sections 2.8.1-2.8.3). Developments in theories of performance and performativity are discussed and fleshed out further with sections on poststructuralism, posthumanism, multimodlity and transmodalities (in Sections 2.8.4 and 2.8.5). I then discuss traditional and alternative approaches to the *construct* in assessment (Section 2.9). In exploring the role of the construct in performance assessments and EAP assessments, I pose questions as to its relevance in particular assessment contexts. Evolution in thinking in critical language assessment and ELF receive attention before I consider theory related to developing a construct specifically in EAP assessment. In Section 2.10, I address the debate concerning the place of the measurement paradigm and the concept of validity in CBA. I trace developments in early validity theory borne out of standardized large-scale assessment, and then consider the relevance of this to classroom-based AOP assessments. I specifically discuss how characteristics of AOP classroom-based assessment determine applications of validity. I finish this chapter with a discussion of the power of tests to do good and harm, glossing important terms of fairness and justice (in Section 2.11).

2.2   The Concept of Oracy

Oracy is defined by Wilkinson (1965, p.13; cited in Doherty et al., 2011) as 'the ability to use the oral skills of speaking and listening' in comparison to literacy, which refers to the skills of reading and writing. Wilkinson's definition of oracy focuses on linguistic skills in verbal interaction. Such a conception of oracy centres on linguistic skills. I, instead, wish to use a broader notion of oracy which moves beyond the linguistic and can be applied to multimodal AOP events and classroom-based AOP assessment practices. My use of the term oracy captures an array of linguistic and non-linguistic skills. During analysis of the data, I decided to draw on Mercer et al.'s (2017) Oracy Skills Framework to facilitate the creation of my conceptualisation of what oracy skills are involved in AOP tasks.
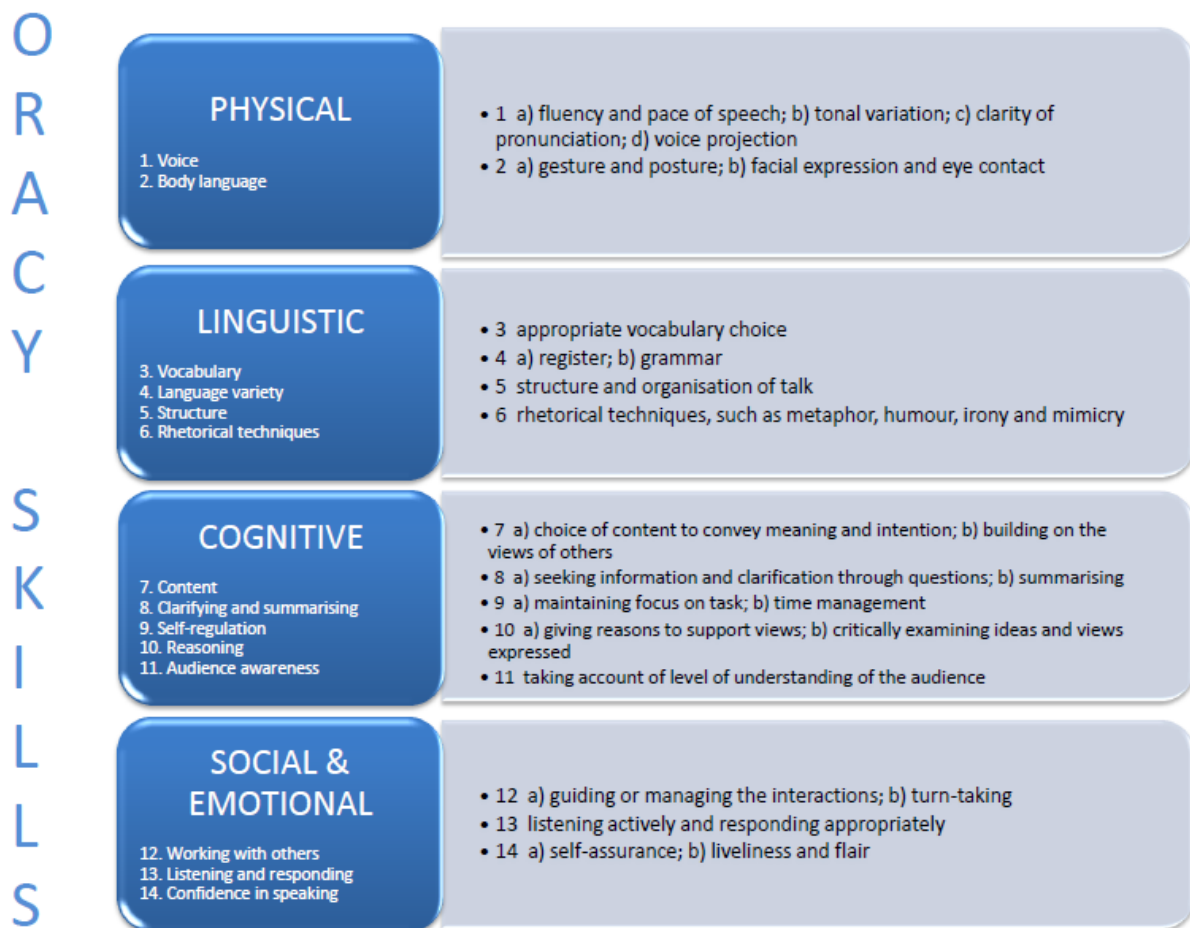
Figure 3: Oracy Skills Framework (Mercer et al., 2017)

The broad range of skills from physical, linguistic, cognitive, to social emotional accounted for a vast amount of codes generated from the data in this study. However, the oral presentation is a blended genre and investigating the EAP AOP construct involved moving beyond the remit, and

expanding the notion, of oracy represented in the Mercer et al.'s Oracy Skills Framework (in Figure 3 above). There are two central ways in which I believe this framework requires development in order to apply to genres which involve an oral dimension in UK tertiary education: its capture of academic skills, and in how it reflects the multimodal nature of much oral communication. In particular, additions to the Oracy Skills Framework which reflect skills often prioritised in HE, such as academic integrity concerns would be welcomed. I believe the Oracy Skills Framework would also benefit from including a section, or parts of sections, dedicated to the use of visuals.

The Oracy Skills Framework was developed based on feedback from practitioners and professionals in secondary schools and on the basis that oracy is 'the development of young peoples' skills in using their first language, or the official / educational language of their country, to communicate across a range of social settings' (Mercer et al., 2017, p. 52). Dippold et al. (2019) state that their research, on ELF communication in group work tasks in UK HEIs, has indicated no problem with using this framework with L1 and L2 speakers of English. I also found the Oracy Skills Framework applicable to the data in this study which centres on AOP performances delivered by L2 users of English. However, the framework and accompanying glossary does not clarify what 'appropriate' would constitute and would benefit from a rigorous analysis as to its relevance in ELF contexts.

In summary, a framework specific to tertiary education could be developed and prove a useful tool for some HE practitioners and students. Despite these limitations, Mercer et al.'s Oracy Skills Framework has proven useful in mapping out broad categories of the skills required in AOP events but further categories have emerged in the data related to learning and assessment purposes across field sites (e.g. formative learning skills: see Section 6.7.3.4) . An oracy framework intended for use in secondary contexts was used due to a paucity of frameworks related to oracy in tertiary settings.

2.3 Oracy Research in HE

This study investigates teachers' and students' assessment practice within HE environments which place increasing emphasis on active learning and group learning (Macfarlane, 2016). The active approach involves students visibly engaging in public spaces, with in some cases such engagement involving oral participation. Consequently, work on oracy in UK Higher Education has received renewed attention in recent years. The inaugural conference specifically devoted to oracy in Higher Education was held at the University of Surrey in 2018. Studies conducted in the Australian HE context (namely Doherty et. al, 2011) and UK secondary context (notably the work of Neil

Mercer at Cambridge University) has served to guide and inform recent studies (e.g. Heron, 2019) on oracy conducted in UK tertiary education.

Albeit a growing area in UK Higher Education, research into oral and blended academic discourse genres in tertiary contexts is still underdeveloped, especially when compared to research on academic written genres. Key studies which have sparked the investigation of oral/aural academic discourse expectations and practices in tertiary contexts include Ferris and Tag (1996a; 1996b) and Ferris (1998). While these studies provided a typology of aural/oral activities in US universities, and revealed the importance of considering differing student and lecturer conceptualisations, they lacked sensitivity to contextual factors. In Ferris' (1998) study, lecturer and student conceptions of expectations and challenges in oral tasks were compared in relation to different contexts. The researchers emphasised the importance of conducting future studies into specific genres. In particular, they called for gaining teacher and student perspectives in the same setting using qualitative research methods.

The prevalence of oral presentation assessment has been addressed in a selection of research studies. In 2007, Joughin noted that oral assessment was growing in use. Hounsell et al. (2007) conducted a review of published papers on innovative assessment studies conducted in UK HE settings or written by UK-based authors. They found that oral presentations feature fairly prominently (in 76 out of 317 papers) in the innovative assessment literature. More recently, Huxham, Campbell and Westwood (2012) highlight that the oral presentation is regarded as the most common type of oral assessment in UK HE (Huxham, Campbell and Westwood, 2012, p.125). Further research which builds a clearer picture of the ubiquity of oral presentation assessment on degree programmes in UK HEIs would be welcomed.

Research has been conducted specific to the AOP genre in tertiary contexts on disciplinary programmes and EAP programmes. These studies have approached the use of an AOP in educational contexts as a learning tool and/or an assessment tool. Studies report the knowledge and skills required in AOPs including linguistic aspects (Zareva, 2009) role of multimedia in improving performance (Zareva, 2011; Hill and Storey, 2003) and skills needed at discourse level (Morrell, 2015; Bankowski, 2010). Past literature has also focused on validity and reliability concerns such as peer and teacher rater reliability (Penny et al., 2005; Langan et al., 2008).

There has been little research conducted which provides a close understanding of the AOP assessment processes, constructs, task functions, purposes, and oracy demands placed on students and teachers at UK universities. Heron's (2019) study begins to address this gap, by conducting case studies on AOP assessments on two business modules at a UK university to investigate the oracy demands placed on teachers and students. She finds that oracy may be a

process of the assessment - needed to complete tasks but not explicitly supported and developed. Alternatively, oracy can be both process and product - needed in the task and explicitly taught. Heron makes the case to embed oracy skills into the curriculum. In addition, Heron calls for further research which develops a closer understanding of what oracy skills are involved in Higher Education tasks. The perceptions of the role of oracy skills in HE is an area which warrants further exploration. In the first strand of this research, I investigate AOP assessment purposes, functions, and treatment of communication on individual taught modules. I examine the oracy expectations of staff who act as assessors on AOP tasks at 7 UK universities (see Section 4.3 for relevant data analysis).

A British-council funded project which investigates oracy skills development on pre-sessional courses and transition to disciplinary studies is being conducted currently. It is a welcome addition to research on oracy. Studies which cover multiple task types and notions such as group work or oracy skills may not appreciate the relevance of genre-specific and task-specific considerations, however. Dippold et al. (2019) argue that 'speaking instruction' in UK HEIs usually covers 'monologic events such as presentations which do not give rise to the interpersonal and language issues' which they had described in their research on group work. However, the oral presentation may involve the dialogic aspects and interpersonal elements they describe (such as negotiating misunderstanding in meaning, and regulating turn taking) most obviously in a question and answer session and in preparation stages in group AOPs. There are few studies which have investigated 'interpersonal and language issues' in delivering and assessing AOPs in UK HE. This study shows how language and interpersonal factors (among others) shape AOP practices (see Chapters 5, 6, 7 and 8 for analysis and discussion of such factors).

The main focus of this research is on what is involved in EAP AOP practices. Due to a focus on authenticity in EAP (Hyland and Shaw, 2016; Harwood and Petrić, 2011), practices in the target domain are of significance. There have been a number of pivotal studies on AOPs in Anglophone EMI settings which have focused on implications for EAP instruction. I have selected three to describe in detail (see Table 2 below). The studies in Table 2 proved particularly instructive in the early stages of this project because they showed ways in which to investigate practices around a central AOP task using qualitative approaches with the aim to inform EAP practice and policy. These studies have been summarised in the table below:

Table 2: Key Oracy Studies

Focus	Author	Methods and relevant findings
Discourse socialization of non-native English speaking graduate students through AOP events	Zappa-Hollmann (2007a and 2007b)	Zappa-Hollmann's (2007a) report on part of her master's level dissertation (2007b) explored discourse socialization of six non-native English speaking graduate students in one Canadian university on different subject programmes through their delivery of an oral presentation. Zappa-Hollmann conducted a qualitative multiple-case study and spent four months in the field. AOP observations, interviews, field notes, course outlines were collected for thematic analysis. Zappa-Hollman's study benefited from framing her report using language socialization theory and an ethnographic methodological approach. The key contributions of this research included a documentation of the main AOP activity characteristics in four disciplines; comparison of student and lecturer conceptions in one context; an acknowledgment of participants' critical orientation towards AOP expectations; a consideration of the type of feedback given on AOPs; and commentary on the nature of obstacles faced by students.
Oracy demands in AOP and group project assessments on subject programmes with interest in implications for English as an additional language (EAL) students	Doherty et al. (2011)	The research was conducted on two first-year Australian university courses with explicit industry orientations and which had a large number of international students enrolled on them. The study uses observational data from video-recordings of lectures, tutorials and oral presentation sessions, course documents and semi-structured interviews including 'stimulated recall' (Keith 1988) with the course coordinators. Doherty et al. (2011) explore the pedagogical designs taken on oracy and the oracy demands of group project and group AOP assessment tasks. The key finding of the research is that the linguistic proficiency required by any curriculum is in danger of being 'implicit' or 'tacit' in pedagogy. The two courses followed in the research approached students' oracy differently: as product or process. The treatment of oracy was shown to have implications for EAL student needs, and the perceptions about the importance of talk and the 'hidden assessment' of oracy.

Focus	Author	Methods and relevant findings
Subject lecturer perspectives on AOPs and EAP presentation textbook evaluation	Levrai and Bolster (2015)	In the first phase of the study, 8 interviews were video recorded with subject lecturers and presentations were observed at University of Nottingham Ningbo China in order to develop a clearer conceptualization of academic oral presentation expectations and features. In the second phases, lecturer expectation findings were then used as a framework to evaluate the effectiveness of current presentation skills textbook training materials targeted at non-native English speakers. Key findings from the first phase of the study: *1. Why AOPs are used in Higher Education:* a. provides important transferable communication skills b. allow students to make their own discoveries and learn independently c. How much and how well students know the subject matter can be identified in academic presentations *2. Features that make a presentation suitable for academic contexts are that it:* a. Contains references and is verifiable b. Contains citations to sources which are synthesized c. Contains solid arguments borne of research *3. Four key features of good presentation:* a. timing b. structure c. audience awareness d. research basis (pp.66-67) Key findings from phase two (textbook evaluation) were that current training materials could do more to develop the specific skills needed to deliver successful academic presentations including more emphasis on the structure and organization of different genres of academic presentation and advice on integrating sources and referencing.

These studies, in particular, have informed the methodological approach I have taken in this study, namely the ethnographically-oriented, qualitative approach to researching AOP practices on subject courses and EAP courses. There is a notable lack of research in the literature, however, on factors which shape student and teacher decision making in assessment contexts and particularly on EAP courses. Work in the field of CBA has investigated assessment processes and decision making (examples provided in Section 2.10.2). While there have been studies which are beginning to develop a closer understanding of AOP practices on taught modules in tertiary contexts, a sufficiently comprehensive picture of patterns and diversity of practices has yet to be painted. Research into EAP AOP practices is lacking (details on this research gap are provided in Section 2.9.3). These research gaps have ramifications for the field of EAP. I now turn to a review of literature which probes the concept of EAP.

2.4   English for Academic Purposes

Work in Applied Linguistics in the area of study of *English for Academic Purposes* has sought to, as Hyland and Shaw (2016) put it, 'demystify' academic discourse. The field of EAP has been defined by many EAP scholars. The most cited definition of EAP, according to Bruce (2017), is Flowerdew and Peacock's (2001) definition:

> the teaching of English with the specific aim of helping learners to study, conduct research or teach in that language. (p. 8)

This definition is described by Bruce (2017) as focusing on the end product of EAP rather than the process. Flowerdew and Peacock's definition also presents EAP's primary purpose as English language development. I do not align myself with this definition as it does not reflect the activities of EAP practitioners as I see them. A number of later definitions have, instead, nestled EAP within a wider context of communication.

> concerned with those communication skills in English which are required for study purposes in formal education systems. (Jordan, 1997, p. 1)

> covers language research and instruction that focuses on the communicative needs and practices of individuals working in academic contexts (Hyland and Shaw, 2016, p.1).

Hyland and Shaw's (2016) definition importantly recognises the status of EAP as an academic research field. Another welcome addition is that of communicative *practices* as this encapsulates aspects of EAP which go beyond language and recognise individuals' experiences. I sought a description which encompassed the roles of language, communication, discourse and grasped the focus on academic and social and emotional skills which I have found to pervade EAP practice and

eclipse focus on linguistic skills in instruction. Hyland's (2018) article captures the role of language, communication and discourse in EAP:

> EAP is an approach to language education based on identifying the specific language features, discourse practices, and communicative skills of target academic groups, and which recognizes the subject-matter needs and expertise of learners (Hyland, 2006). It sees itself as sensitive to contexts of discourse and action, and seeks to develop research-based pedagogies to assist study, research or publication in English… EAP practitioners employ different streams of theory and research in their work. Central to this understanding is a focus on *discourse* rather than just *language* and how communication is embedded in social practices, disciplinary epistemologies and ideological beliefs.
>
> (p.2 and p.12; italics in original)

Hyland's description comprehensively addresses the concerns which EAP stakeholders confront, especially in stressing EAP's focus on 'discourse rather than *just* language' (my italics). This represents the development of the conceptualisation of the field of EAP over its lifespan from a focus on English language to include a broader focus on language, communication and discourse, alongside recognising how subject literacy and philosophical approaches shape practice.

EAP as a field has encountered, and remains to face, significant issues. Work in Critical EAP, Academic Literacies and English as a Lingua Franca (in Academic Settings) (henceforth ELF(A)), (e.g. Benesch, 2001; Lillis and Tuck, 2016; Leung, Lewkowicz and Jenkins, 2016) has presented EAP as static and normative in its approach, emphasising the need for teachers and students to consider how epistemological orientations, ideology and power relations at discipline, modular and individual level shape which ways of communicating accrue cultural capital in the academy.Ken Hyland and Philip Shaw, key figures in EAP, have acknowledged criticisms posited by critical EAP, ELFA and Academic Literacies scholars, for instance, stating that EAP provision has potentially put too much emphasis on supporting second language users into their disciplinary environments (2016, p.11). An increasing number of courses are now offered to 'home' as well as 'international' students, who may be first or second language users of English. This research has taken place on programmes with L2 speakers of English enrolled on them, however, which remains the overwhelming trend. The justice of such practice is an important concern – this is not the primary focus of my research but is addressed in Section 9.4 on implications.

The constraining social structures which shape the status of EAP units themselves in institutions and the limiting effects this may have on practices have also been exposed (most notably in Hadley, 2015 and Ding and Bruce, 2017). EAP teachers at UK universities are often 'vulnerable, short-term instructors in marginalized 'service units'' (Hyland and Shaw, 2016, p.6), and EAP units now deal with increasingly greater competition from private providers. Ding (2016) describes this period as a 'watershed moment in EAP' in which EAP practitioners need to address their

somewhat precarious position in the academy and how they can specify and promote their role and identity in the wider academic community.

The discussion of EAP and the definitions of EAP often gravitate towards literacy (centred on reading and writing linguistic skills among others). This has left oracy somewhat marginalised in the field. EAP instruction often covers preparing students for 'seminars' and 'oral presentations' (Schmidt, 2015), however the research base dedicated to investigating such genres is lacking. This is surprising as a genre approach has been identified as a trend in EAP scholarship (Benesch, 2001) and from my experience a genre approach underpins curriculum design in some institutions. The next section discusses Genre Analysis.

2.5   Genre Analysis

Benesch (2001) identifies genre analysis as a trend in EAP research. Approaches to genre analysis research include moves analysis and ethnographic approaches. Move analysis involves identifying sections of texts which perform specific communicative functions (moves), whereas ethnographic approaches include methods such as conducting interviews with the producers of the texts; a mixture of textual and ethnographic approaches is often termed a textographic approach (Swales, 1998). Swales (2019) acknowledges the limitations of move-step accounts and calls for ethnographic methods to be used:

> So when I read unadorned move-step accounts of emerging or relatively recent genres, such as elevator-pitches, graphical abstracts, three-minute theses, and conference posters I find them insufficiently "thick" in Geertz's sense (Geertz, 1973). Further, the structures revealed often seem to be highly predictable. If, for instance, essentially the same structure occurs in conference posters and in journal articles, or in conference presentations and 3M theses, then we need to look at other aspects for those differences—differences that can then be deployed in our efforts to prepare our students for these genres. And this will likely involve going beyond the textual or transcriptual to interviewing producers and recipients, along with those who have gatekeeping or evaluating roles. (Swales, 2019, p.77)

In this study, I seek to develop a closer understanding of AOPs used as assessment tasks on taught degree modules and on EAP modules. The prime focus is on EAP AOPs, therefore the fieldwork using ethnographic methods takes place in EAP settings only. Such fieldwork can offer the thick description Swales calls for. The survey study of taught degree module and EAP module AOP practices teamed with the fieldwork in EAP settings develop a closer understanding of AOP assessment tasks, which contributes to our notions of an AOP. Not only how to research genres, but approaches to the concept of genre itself varies (discussed in Section 2.6 below) and this research also contributes to such debates. In the fieldwork study, I access what is valued in AOP assessments by doing what Swales advocates: I not only interview, but also observe, 'producers',

recipients' and those 'who have gatekeeping and evaluating roles' (I elaborate on the methodological approach in Chapter 3). Swales (2019) also argues that 'the scope of a genre can be reconsidered in order to serve the pedagogical purposes of the EAP practitioner' (2019, p.76). The scope of the genre can also be reconsidered for research purposes. I focus on EAP AOP assessment as I seek to understand how the assessments are interpreted and what this may tell us about the construct of EAP and gatekeeping decision making.

Knowledge of genres makes the potential boundaries and creative opportunities visible to EAP stakeholders (Hyland, 2018). Research and instruction which approaches particular genres, must take into account the complexity, diversity and indeterminate nature of genre (see Section 2.6). Ethnographic studies by EAP scholars (e.g. Prior 1998) and corpus studies (e.g. Nesi and Gardner, 2012) convey an image of notable diversities and different functions across and within academic written genres. It is essential to note that the use of exemplars of genres does not entail promoting conformity.

In EAP, the need to focus on 'processes and contexts' more than texts has been voiced (Belcher 2006). Ethnographic studies are advocated by Academic Literacies scholars who seek to focus on the processes which create the products (Connor, 2004). Academic Literacy proponents have underscored the 'cultural and contextual component of writing and reading practices' and their effects on students' learning (Lea and Street, 1998, p. 158). This quotation highlights that Academic Literacies research is predominantly centred on written texts and practices, although work on multimodal practices has also been conducted under the umbrella of Academic Literacies. EAP practitioners also draw on perspectives less traditionally found in EAP. The ecological approach to language (Hayward, 1995) for instance has informed EAP courses (e.g. Garner and Borg, 2005), and inspires the teacher Adam in FS1 in this study. Approaching practice from a genre standpoint can complement such other approaches to teaching, learning and researching EAP.

In the next section, the concept of genre is unpacked and previous attempts at describing oral presentations are critically evaluated.

2.6   Genre and AOP

In this research I explore academic oral presentations. This term is conceptualised as a genre in this study. Debate ensues as to what a genre is, whether it exists, how it is defined and what denotes belonging to a given genre. Traditional definitions of genres connote conventions of content, and/or form shared by texts belonging to them. Within a genre, texts contain unique features which will break the bounds of such definitions, however. Miller's (1984) seminal work

reconceptualised genres as social constructs and actions. Miller (1984) notes what may be perceived as similar tasks and communicative purposes may also diverge in fundamental ways. Devitt (1993, p.575-6) notes that genres involve purposes, participants and themes and therefore 'understanding genre entails understanding rhetorical and semiotic situation and a social context'. A genre is an abstract conception (Feuer, 1992, p.144; cited in Chandler, 1997, p.1). It is nonetheless a useful concept which represents the relationship between text and context. While on an intuitive level grouping and recognising texts as belonging to particular genres can aid communication, on closer inspection such categorisation is problematic. Genres are social-cultural constructs which describe social conventions, but these constructs and their conventions are realized in variable ways by individuals (Widdowson, 2020, pp.91-95).

A number of genre theorists have conceptualised genres using Wittgenstein's (1958) concept of family resemblances rather than through definitions. The term family resemblance connotes an argument that not all entities in a group share the same features as all other entities: members in a grouping may share overlapping similarities but any two instances may not have a common feature. Swales (1990) regards such an approach, in line with Vygotsky, as 'undisciplined chaining' and even describes it as irresponsible. But it is not without some merit, according to Swales. He believes that a 'privileged property' should be identified, which in his case is *communicative purpose* (1990, p.58). However, he accepts that the family resemblance concept is useful as not all members of the genre may share this 'privileged property'. Swales aligns himself with Armstrong et al.'s statement:

> There are privileged properties, manifest in most or even all examples of the category; these could even be necessary properties. Even so these privileged properties are insufficient for picking out all and only the class members, and hence a family resemblance description is still required. (1983, p.270; cited in Swales 1990, p.52)

From encountering different AOP texts and descriptions of AOP formats in this study, I have considered what the privileged properties may be and found the concept of family resemblance to reflect the realities of diverse real-world assembled communicative acts which are conceptualised as AOPs. I endeavour in this thesis to avoid characterising or reifying AOPs as fixed and stable entities. The formats, purposes, characteristics and practices associated with the tasks differ significantly according to the people and settings. I have needed to establish some degree of shared understanding with informants in the research about what is an AOP task and have reflected on the negotiation of the term academic oral presentation (see Section 8.2 for a discussion of such negotiation). The genre can act as a useful anchor but it is an ephemeral and shifting one due to individual applications.

Widdowson (2020) describes how genres can be positive force in creating shared expectations which facilitate cooperative communication:

> … generic conventions or norms of communicative behaviour effectively stake out the spaces of shared social territory and set the limits within which individuals are expected to co-operate. So at a social level, the common ground for interpersonal positioning is already in place at the onset of the discourse the participants are to engage in. (Widdowson, 2020, p.90)

Individual's lack of receptiveness and/or willingness to conform with pre-defined generic conceptions determine the realizations of genre. The emerging navigation of communicative practice may diverge from expected behaviour and in some cases create new conceptualisations of generic conventions. Or, the genre may be used as a negative force by limiting justified variation by individuals (Widdowson, 2020, p.95). This negotiation of genre conventions is important to consider in assessment contexts.

In previous research which has investigated academic oral presentations, researchers have offered working definitions or descriptions of the genre. The term 'oral monologue' or reference to 'conference presentation' or 'oral presentation in academic settings' has been used in some instances instead of 'academic oral presentation'. Ming (2005) describes AOPs as a 'partly spoken and partly visual form of communication' (p.118) which usually has a time limit and takes place in organizational settings. Hyland (2009) gives insights into the oral presentation by comparing them to lectures and seminars. He likens AOPs to lectures in that they are a monologic discourse that deals with information transfer, and he finds similarities with seminars in their informal and conversational aspects. Few researchers have grappled with what features of an oral presentation make it academic. Levrai and Bolster's (2015) study (see Section 2.3, Table 2 above) proposed three such features:

a. Contains references and is verifiable

b. Contains citations to sources which are synthesized

c. Contains solid arguments borne of research   (p.67)

Levrai and Bolster's work proves helpful in developing a closer understanding of the genre. However, there are presentations which do not contain verifiable data, have no citations and contain no arguments from research, which would be deemed as academic. I attended a conference presentation recently in which a renowned academic in the field of Applied Linguistics spoke about fictional data and considered what resonance this may have to a particular field. It contributed to a discussion, which was borne out of previous research but did not *contain* such features that Levrai and Bolster describe. This illustrates the unenviable, or unachievable, task of any such contained description of an AOP.

While offering a description of conference presentations, Swales (2019) argues that genre analysts and EAP practitioners may approach description of a conference presentation in a different way:

> Or consider the case of academic speech, particularly the very common speech event of a conference presentation (often 20 min for the presentation itself, followed by ten minutes for discussion or Q and A). As genre analysts, we will probably want to conceive of both parts as constituting a single speech event. We recognize the presentation consists of a single main speaker, the same participants in the same room at the same time with everybody's attention focused throughout on the main topic (or so the presenter hopes). However, if we are being EAP practitioners preparing our students for this particular experience, we will (obviously?) opt for two genres because the cognitive loads, emotional tensions, and linguistic and rhetorical expectations and affordances of the two parts are very different. In this situation we recognize that the first is basically monologue, while the second is multilogue. Part 1 is prepared (and probably rehearsed); Part 2 is unscripted (and probably in parts unexpected). Finally, the language and paralanguage requirements of the two sections have different exigencies and affordances. In the first, speakers address the whole room; in the second, they serially direct their attention to the various individual questioners and commentators. (p.76)

The description above closely mirrors many, and significantly departs from a few, conference presentations I have attended. Swales argues that 'the scope of a genre can be reconsidered in order to serve the pedagogical purposes of the EAP practitioner' (2019, p.76). This extends to practitioners in different contexts who use AOP tasks as pedagogical and assessment tools. In this study, I am investigating the academic oral presentation tasks which act as assessment tools on UK Higher Education taught courses. In this study 'academic' is used to denote those which take place in academic settings on university taught programmes and EAP courses in tertiary education institutions. I, therefore, as a researcher view the scope of the genre to serve the investigative purposes of this study. In other words, my interpretations of these events are shaped by primarily categorizing these AOPs as assessment tasks. The research purpose itself determines the concept of AOP genre portrayed in the study. I nevertheless endeavour to reflect informants' views on what the AOP involves in their contexts even if the research perspective may distort this to some degree. I offer musings on the construct of an AOP genre in 8.2.

The next part of this chapter (Sections 2.7 and 2.8) reviews literature on assessment in educational settings and on second language performance assessment. EAP assessment has been largely influenced by developments in language testing. The following section provides an account of influential trends in language testing and assessment and points to how trends and concepts are useful in understanding AOP practice in UK HE.

2.7   Trends in Language Assessment

The assessment practices in the EAP settings explored in this thesis contain traces of numerous discourses in language and educational testing and assessment. In this section, I set out the historical trends in language testing and assessment, drawing particularly on trends identified in Anthony Green's (2014) book, which are particularly relevant to, and detectable in, EAP AOP assessment practices investigated in this study.

Up until the 1970s, language was viewed as an abstract system. This was reflected in approaches in language testing, which used discrete-item tests where linguistic components were tested separately. The discrete-skills test often focused on components. Lado's *skills/components* model, for instance, classed speaking as a skill and pronunciation as one of the components. Carrol in 1961 advocated a skills/component approach which prioritised skills. Carrol introduced an *integrated* approach focused on a language user's ability to produce or comprehend complete utterances. Campbell and Wales (1970,p.247;cited in Green 2014, p.198) identified the important ability of understanding or producing utterances which are not only grammatical, which had been the focus previously, but also how 'appropriate to the context' the utterances were. In the 1970s, communicative language testing was therefore implemented due to this shift in perspective, following work in sociolinguistics (notably the work of Hymes).

A task-based communicative approach in which the ability to use language functionally to perform tasks was to be assessed was advocated and took sway in language testing. However, limitations with the communicative, task-based approach were identified. Bachman (1990) argued that a definition of skills was needed and completing a 'real-life task' was no substitute for this. Consequently, Bachman (1990) and Bachman and Palmer (1996, 2010) integrated the skills/components tradition with the communicative approach. Language knowledge is included as well as characteristics such as strategic competence and affect in this model. Bachman and Palmer regarded the model's components as distinguishable from other 'attributes of the individual language user' (e.g. personality) and characteristics of the language use situation and other language users (2010, p.300). Bachman's models were critiqued by McNamara (1995, cited in Green 2014) who believed testing individual abilities ignored the co-constructed nature of interactive tasks. An interactional view gained traction in which contexts and fellow language users become integral to language ability (Chapelle, 1998; Chalhoub-Deville, 2003). Language became viewed as inseparable from context and learning as a social process with the learner socialised into local cultural practices. Frameworks and scales therefore were viewed as reflecting dominant value systems which are open to shift. Socio-cultural theorists view assessment

constructs not as fixed properties, internal to the human mind, but unstable ones governed by particular social values. This implies that an assessment therefore becomes bound to the local and the ability to extrapolate the inferences made by test scores to other settings and tasks is problematic.

The view of assessment's role in learning has also shaped trends in language assessment. The notion that education involved mastering a fixed body of knowledge was challenged and formative testing has since been championed. Students would need to adapt to changing demands in life and the learner's ability to direct their own learning was considered crucial. Assessments came to be viewed as not only tools which measure learning, but a tool that can support learning. The concepts of formative (or *Assessment for learning*) and summative (or *Assessment of Learning*) assessment have been much discussed in assessment literature. Since Bloom's application of the concepts to the teaching-learning process (from Scriven's coinage of the terms in work on programme evaluation), there has been a lack of consensus (especially a divide between test developer and teachers) on the artifacts or practices which connote formative and summative assessment (Dunn and Mulvenon, 2009; Bennett, 2011). In this thesis, I use the concepts of 'formative' and 'summative' to describe aspects of assessment practice. The formative factor, as I describe it, in assessment practice relates to actions which are primarily targeted at improving learning and teaching; whereas the summative factor describes aspects of practice which are predominantly focused on judging what has been learned or achieved, or focused on an assessment score outcome rather than long-term improvement. There is much overlap between the two concepts however, and the practice in the field sites in this study is a complex blend of formative and summative approaches (see evidence of this in Sections 6.7.3, 7.8, 7.9 in particular). Learner-centred approaches have been adopted in language classrooms through the concept of *Assessment for Learning*. Learners also participate in evaluating performance according to this theory. Peer assessment and self-assessment became part of language assessment practices in language classrooms in the 1990s.

The historical developments in language testing are relevant as it serves as a vital backdrop to understanding the complex, hybrid nature of AOP practices on EAP courses. The AOP assessments which feature in this thesis contain traces of discrete skills testing, in that teacher assessors and the criteria separate language (and other skills) into components such as pronunciation and grammar. The influence of *Assessment for Learning* is visible in particular emphasis on the formative functions of the AOP tasks in FS1 and FS2 (demonstrated in analysis Sections 6.7.3 and 7.9). The interactional view and need to consider the social dimension of assessment is traceable in this investigation of AOP assessment practices. This study adopts a social orientation

(advocated by McNamara 1997) by considering the interaction between assessors and assessees during and prior to the AOP assessment delivery (for examples of this see analysis Sections 5.6 and 6.8 and 7.6). The task-based communicative movement is the trend in the rich history of language assessment which is arguably the most visible and impactful on the AOP practices. The next section is devoted to a discussion of performance assessment as theories in this area have particular resonance in this project.

2.8   Second Language Performance Assessment

Real-life performance is increasingly the criterion of choice for language testers and integrated tests, specifically performance-based ones as a result are used more frequently. Performance assessments are described by McNamara:

> A defining characteristic is that actual performances of relevant tasks are required of candidates, rather than the more abstract demonstration of knowledge, often by means of pen-and-paper tests.
> (1996, p.6)

A performance-based approach was initiated by Hymes' (1972) work and gained impetus and justification in the advent and acceptance of the theory of communicative competence. Language performance tests were implemented in the main to ensure assessment practices were in sync with the communicative teaching approach being adopted (McNamara, 1996). They were introduced in EAP circles due to a call for integrated tests over discrete tests.

A key distinction in performance testing is that between *test* and *criterion*. Performance tests may aim to simulate the non-test real-life tasks (criterion) but the test situation (test) is distinct from the 'real-life' performances. I adopt a broad notion of performance test which includes tests which 'emulate some aspects of nontest settings' (Haertel, 1992, p.984; cited in McNamara 1996). This is not to say that the whole act of performance assessment is not artificial. It is artificial, but performance testing seeks to bear some resemblance to tasks which are not constructed for formal assessment purposes. I conceptualize the EAP AOPs as performance assessments. There are three crucial issues, with satellite areas of concern, in investigating EAP AOP performance assessments: the role of the criterion, the place of linguistic and non-linguistic factors in the tasks, and how the concept of performance is theorised. The following sections flesh out these key questions which surface throughout the thesis.

2.8.1   The Criterion

The use of AOPs as gateway assessments on EAP courses is widespread, however there has been little empirical research conducted as to what AOPs involve in the criterion, also referred to as the

target domain. This points to a fundamental concern that the assessment task and the criterion may be being conflated (Davies, 1992, p.9). The further away from criterion, the weaker the inferences can be drawn from the performance test that test takers will perform in the real-world tasks. In second language performance tests the distance from the tasks in the target language use domain is arguably in some instances greater, as language is not only the medium but a larger part of the target of the performance, than in the criterion. McNamara's (1996) called for an improved understanding of what performance involves and the role of underlying performance abilities in second language performance assessment.

> Progress in understanding the nature of second language performance testing and conducting the necessary research on its validity requires a commitment to rigorous analysis of what is involved in a language performance test. (p.88)

Since 1996, there have been a number of studies which have outlined the benefits of performance-based assessment such as ability to tap into complex abilities, greater motivation and flexibility leading to greater equity (e.g. Glisan et al., 2007; Colley 2008). This research aims to develop a closer understanding of what is involved in AOP performance tests in EAP settings. I develop descriptions and conceptualizations of what second language performance in an EAP AOP involves and, in doing so, consider the factors (linguistic and non-linguistic) which shape teacher assessor and student test taker actions and thinking. In the survey study, I gather data on task formats, purposes, aspects assessed and the treatment of oracy skills and begin to consider in what ways gateway AOP tasks on EAP courses emulate AOP tasks in UK Higher Education Institutions (see chapters 4 and 8 for data analysis and discussion).

2.8.2   Linguistic and Non-Linguistic Factors

A crucial issue in performance assessment is the role of linguistic and non-linguistic aspects in a task.

> The distinguishing feature of performance tests… is that they tap both second language ability and the ability to fulfil the nonlinguistic requirements of given tasks (Wesche 1992, p.105 cited in McNamara 1996, p.39)

McNamara poses questions such as: What is the role of language in relation to non-language factors in a second language performance test? (p.8-9). He also identifies 'strong and weak senses' of second language performance tests which depend on the degree to which assessment criteria include non-linguistic aspects of the task performance. Weak second language tests consist of assessment criteria heavily linguistic in focus.

In performance testing, the question arises as to whether the performance is primarily the vehicle or the target of the assessment. The same applies to the role of language in an EAP performance

test: is language primarily the vehicle or the target? There has been increasing understanding of the complexity of what second language and first language performances in different domains and settings may involve. Researchers have endeavoured to map how language knowledge interacts with other non-linguistic factors (Goffman, 1967; Hymes, 1972; Munby, 1978; Brown, 1995). The importance of non-linguistic features in shaping performance has been argued. Factors other than language use such as motivation and stage confidence and professional competence shape successful performance (Harding and McNamara, 2018). Ervin Goffman's (1967) seminal work on interaction rituals identifies qualities which influence performance such as 'courage, gameness, gallantry, composure, presence of mind, dignity, stage confidence' (p.224). The non-linguistic factors deemed necessary in AOP performances, including what constitutes professional competence or 'an ability to cope in academia', is investigated in this thesis (these qualities pervade the data on EAP assessment in Chapters 5-7). Larger spread of scores for native speakers on IELTS compared to TOEFL was considered to be due to the stronger performance orientation, with larger degree of importance of non-language specific variables, of IELTS than in TOEFL (Hamilton et al., 1993). It has been argued that examinations of native speaker performances in performance assessments demonstrate the 'untypical and elusive' nature of an idealized native speaker performance and therefore its flaws as a reference point (Hamilton et. al 1993; See Section 2.9.2 for further discussion on the relevance of native speaker as a benchmark). In addition, the ability to separate out linguistic and non-linguistic factors alongside each other when assessing AOPs in real-time is problematized (see Sections 4.4.5 and 6.3 for analysis and 8.5 and 9.2.1 for discussion on this phenomenon).

The conceptualisations of knowledge, competence and performance, and the interaction between them, are fundamental to theorizing how a performance is interpreted and comes to be. The next section grapples with the concept of performance.

2.8.3   The Concept of Performance

The notion of performance and performativity have become central to linguistics and many other fields. Scholars have argued for a shift in focus away from competence to performance (e.g. Bourdieu, 1991; Pennycook, 2004; Canagarajah, 2006).

Noam Chomsky introduced a distinction between competence and performance. The distinction and relationship between competence and performance has since been challenged and reconceptualised by numerous scholars. Hymes (1972) differentiated between knowledge, ability for use, and actual use and actual events. In his *ability for use*, Hymes posited a shift from how one speaks of the performance of a motor and that of performance of a person.

> …the performance of a person is not identical with a behavioural record, or with the imperfect or partial realization of individual competence. It takes into account the interaction between competence (knowledge, ability for use), the competence of others, and the cybernetic and emergent properties of events themselves. A performance, as an event, may have properties patterns and dynamics not reducible to terms of individual or standardized competence. (p.283)

Hymes draws attention to the idea that a performance is not wholly reducible to an individual's partial realization of pre-existing competence. The competence/performance distinction, as outlined by Chomsky, is also problematized by Canagarajah, as it assumes that cognitive abilities primarily shape practices. For Canagarajah, competence can be a physical activity, competence shaped in and by the actor's environment, what Hymes describes as 'cybernetic and emergent properties of events themselves'. Instead of competence, Canagarajah (2018) conceives of competence rather as 'emplacement', a term he uses to index 'the possibility that communication and cognition are shaped by the agency of material and spatial features.'(p.50). He argues that there are instances when a resource of the repertoire will come into being, potentially co-constructed with others: a communicative performance is 'assembled in situ' as Canagarajah describes it.

The idea that a performance does not result from standardized individual and prior competence and comes into being suggests that the communicative acts being described are *performative* in nature. Austin (1962) coined the notion of performativity and argued that meaning was understood not to be drawn from symbols but from the social context and purpose for which they are used. Judith Butler's (1990) work on gender developed and popularized the concept of performativity in a range of disciplines. Butler conceptualised gender as continuous performances instead of a representation of a prior identity. Pennycook (2004) made the case for the concept of performativity to be taken up in language studies, arguing that it opened new avenues for conceptualising languages as entities, language and identity and language as part of transmodal communication (see Section 2.8.5 below on multimodal and transmodal communication).

While the performative, emergent, dynamic nature of language and communication has long been noted (e.g. Labov, 1970; Pennycook, 2004; Baird et al, 2014, Baker, 2015; among others), the notion of competence preceding performance retains relevance for some scholars. According

to Widdowson (2020), for a communicative act to take place in English there is often some degree of prior abstract linguacultural knowledge to refer to, build on and depart from all tied together by an abstract system 'the virtual language', which resides in an individual. Those taking an emergent position posit that competence does not necessarily prefigure performance (e.g. Pennycook, 2004; Baird et al. 2014, Baker, 2015, Canagarajah, 2018).

The performative dimension of language, identity, and other features and factors tied up in interaction adds to the complexity of assessing and researching (AOP) communication. It raises the question as to whether emergent, hybrid performances can be tested or assessed. I take up this issue in Section 2.9 on the construct in language assessment. Before this, I consider what compelling poststructuralist and posthumanist theorisations of what is happening in communication offer to the study of AOP phenomena.

2.8.4   Poststructuralism and Posthumanism

Applied linguists who adopt a structuralist orientation view language as separate from context and as a closed structure. Poststructualists and those before them (most notably Labov), on the other hand, reject this view in favor of assigning value to all forms of semiotic activity (e.g. Gee, 1996; Pennycook, 2001; 2018) inextricably tied with language in space and time.

> Situating communicative interactions in space and time accommodates diversity and unpredictability. Conceiving of language and other human activities as abstract and autonomous structures, however, tends to favor homogeneity, normativity, and control. Structures are abstracted from the messiness of material life and social practice. In making structures fundamental and generative, structuralism imposes order and control over material life. (Canagarajah, 2018, p.33)

Posthumanist scholars posit criticism at much social science research for being structuralist, anthropocentric (focus on human) and logocentric (focus on human language and text) in nature (Hultman and Lenz Taguchi, 2010). A 'material turn' has gained momentum in recent years (Kuby 2017) and therefore 'more-than human ontologies' are being adopted by those researching literacy and language (Kuby et al., 2015). Posthumanists argue that objects and artefacts can shape human thought/action as well as human thought/action. As Kuby et al. state: 'posthumanist scholars would say their focus of analysis is on the human ←→ nonhuman (materials) entangled together' (p.13) rather than humans influencing materials. Kuby criticizes human ontologies:

> The focus of agency is placed on humans and does not account for the intra-active agency of materials-with-people and people-with-materials.'(Kuby et al., 2017, p.13)

In this research, I consider intra-action between humans and non-human objects (e.g rating scales and PowerPoint slides) as well as between humans. The AOP assessment tasks involve engagement with multiple resources in multiple settings. Therefore, Canagarajah's (2018) spatial approach is a useful frame by which to view AOP practices.

> From a spatial orientation, communicative proficiency involves the ability to align diverse semiotic and spatial resources for successful activity. Along with the flat ontology of assemblage, it holds that all resources have to be brought together for successful communication. Also, beyond giving primacy to the mind, it posits that the body and material objects facilitate thinking. (Canagarajah 2018, p.49)

In a spatial approach, people not only exert agency on objects, but objects have the ability to shape a person's thinking. In the communicative acts investigated in this research the students and teacher assessors have access to a range of semiotic resources. The concepts of multimodality and transmodalities are crucial in understanding decision making in AOP preparation, delivery and reception.

2.8.5   Multimodality and Transmodalities

Multimodality is a core concept in communication and particularly in the AOP assessment practices captured in this study. Jewitt (2017) gives this definition of multimodality:

> Multimodality describes approaches that understand communication and representation to be more than about language, and which attend to the full range of communicational forms people use— image, gesture, gaze, posture and so on—and the relationships between these. (p. 15)

There are a range of communicational forms used in AOP communication to create and share meaning. Kress (2017, p.60) includes layout, writing and moving image in his definition to reflect the use of multimedia and technology in contemporary communication. As I interpret it, while the notion of multimodality highlights the role of diverse mix of forms used, transmodality stresses the process of integration of modes in communication. Hawkins (2018) proposes the notion of transmodalities to recognise 'the fluid integration, and mutual informativity, of repertoires of resources in meaning-making processes across local and global encounters and interactions in our globalized world' (p.56). The trans- prefix describes the entangled processes at work – the modes are used in conjunction and are blended in their use. The presence of modes shape other modes and so they become blended and inseparable in use. In the analysis of multimodal communication, there are choices to be made between labelling modes and analysing their role in communication and investigating the relationships between modes and how they are intertwined. There are playoffs involved in the two approaches: attention paid to modes separately may overemphasize the role of the specific and not reflect the holistic view; focusing on the

multimodal assemblage may overlook the power of particular modes in communication (Hawkins, 2018). In this research, informants often separate out modes to discern the important roles they exert and this is reflected in how the data is presented, particularly in Chapters 6 and 7.

The semiotic style in AOP communication, rather than a focus on written and spoken communication has been noted by researchers (Kress, 2010; Prior, 2005). Nickerson (2013; cited in Barret and Liu, 2016) describes the AOP as a hybrid genre due to its multimodal nature. In their review of research on AOPs, from an EAP and computer-assisted language learning (CALL) perspective, Barrett and Liu (2016), note the lack of research on AOPs and posit that it is an area of future interest to EAP and CALL researchers.

> Although both EAP and CALL research are at onset of paradigm shifts that reconceptualize the way academic English is viewed, there have not been any systematic reviews of EAP oral presentations, despite its growing prevalence in academia internationally. (p.1230)

Barrett and Liu here identify a link between understanding communication in AOPs and conceptualizing academic English. AOPs, with the heightened visibility of often a diverse array of semiotic resources upon which to draw, are arguably ideal sites in which to contribute to debates on communicative practices. I gathered information on the AOP assessment practices from test taker and assessor perspectives as windows on what is done and valued in such communication. An empirical investigation of decision making strategies and valued practices in AOP assessments in UK HE contributes to understandings of the concept of performance, the role of linguistic and non-linguistic factors including non-human artefacts in academic communication delivered through the medium of English.

The reality of how assessments are implemented by teacher assessors and student test takers has ramifications for the assessment format and what abilities should be chosen as the target of the assessment. The following section covers different approaches to the construct in language assessment.

2.9   The Construct in Language Assessment

The argument (in Section 2.8) that language and identity are performed and created in situ poses a challenge to the traditional notions of an assessment construct and validity. The following sections detail differing approaches to construct development before summarizing the critical reappraisals of the concepts of construct and validity.

2.9.1 Approaches to Construct Development

Bachman and Palmer (2010, p.43) define the construct in language assessments as the 'specific definition of an ability that provides the basis for a given assessment or assessment task and for interpreting scores derived from this task'. The view in language testing regards such a term as fluency as a construct only when it is defined in a way that it can be measured. (Fulcher and Davidson, 2007). Lyle Bachman (2007) identifies a steadfast dialectic in language testing research since the 1960s between the focus on language ability and task/context as the construct to be measured. McNamara (1996) identifies two main approaches to second language performance assessment which present Bachman's (2007) dialectic in a different way, specific to McNamara's conception of performance assessment.

1. *Work sample* approach
2. *Cognitive* approach

The work sample approach involves the use of procedures involved in the criterion and it is interested in the performance. The work sample approach may be regarded as more atheoretical, however a careful description of the context, for McNamara, indicates that these tests are often in line with Hyme's work and therefore sociolinguistic in their positioning. In the cognitive approach, the focus is on the quality of the performance and on what the performance demonstrates regarding the 'underlying state of language knowledge' (p.6). The cognitive/psycholinguistic type of test uses the performance as a vehicle (Messick, 1994) in order to access the target of the test which is some underlying knowledge/ability. *Construct validation* is necessary in both approaches according to Messick. While there may be concerns as to the task's representativeness of the criterion domain task(s) in the work sample approach (Linn, Baker and Dunbar, 1991; Messick, 1994), in the construct-centered approach there are concerns as to whether the choice of task fits the target construct.

Early validity theory assumed that constructs exist, exist independently within the minds of the test takers and are measureable. Validity theory has evolved to reveal new conceptions of validity and measurement (see discussion in Section 2.10 below). However, in many assessment settings, including those featured in S2, the traditional notion of a construct continues to permeate perceptions and practices associated with AOP assessment tasks. A particular model of the assessment and design process gains traction (potentially unhelpfully): once a construct is defined, a test is designed as an operationalized representation of this construct. A test-taker's performance on said task is supposedly used to make inferences on how they perform on the pre-specified idealized and abstracted focal ability/abilities. The process often requires raters to seemingly map a performance onto a rating scale. This practice assumes that an ability can be

specified in advance and therefore be independent of context. The inferences, often in the form of score outcomes, are presented as evidence on how the test taker will perform in the target language use domain (Bachman and Palmer, 2010). The following figure from Jenkins and Leung (2019, p.89) represents this process, of which Jenkins and Leung are critical:
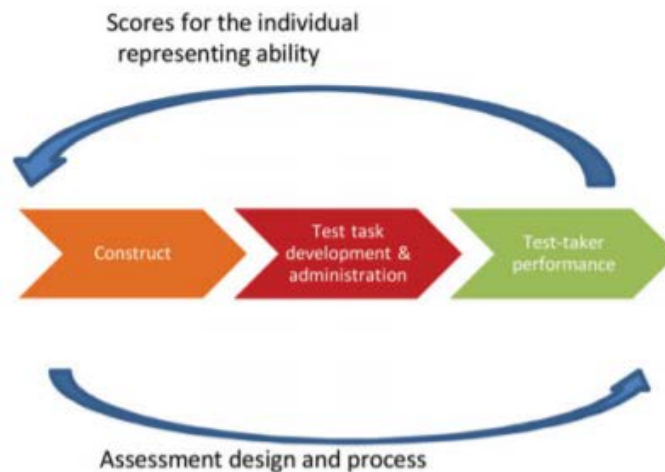
Figure 4: Traditional Relationship between Construct and Performance (Jenkins and Leung, 2019)

Jenkins and Leung (2019) warn of the reification which results from such a process: the construct in its abstraction and pre-specified form has a strong potential to weakly resemble real-world communicative practice. The AOP assessments in both field sites in this study are conceptualised initially by teacher assessors in S2 as following the process depicted in Figure 4 above. However, observations of practices demonstrates that such a fixed, stable and predictable notion of classroom-based AOP assessment process is flawed and not reflected in practice (see discussion on CBA in Section 2.10). The construct evolves and is negotiated (see Sections 5.4, 5.5, 7.2 and 8.4 on this phenomenon in this project). The pervading tension is the need for a degree of shared understanding as to the abilities being assessed and a degree of flexibility to allow a reasonable degree of negotiation in local settings. One way forward is to conduct further and intensified research from stakeholder perspectives which gives indications as to what is valued in criterion and test domains. Constructs are then based on such research and opportunities for justifying movement away from pre-defined constructs are built into assessment processes and openly acknowledged.

The next section traces developments in thinking in critical approaches to language testing and assessment.

2.9.2   Critical Assessment and ELF Perspectives

Shohamy (2018) argues that the current testing theories and tests use outdated language constructs which do not reflect recent research in areas such as translanguaging, ELF, multilingualism and multimodality. Research in critical language testing, translanguaging, multilingualism, multimodality and English as a Lingua Franca (ELF) e.g. Shohamy, 2011 and 2018; Li Wei, 2018; Baird et al., 2014; Jenkins and Leung, 2016 and 2019), promote the reality of language practices as emergent, fluid and interconnected and the assessment of language use in multilingual or ELF settings against a benchmark of fixed, static or norm-based criterion not based in the reality of individuals' repertoires (Blommaert, 2010) or local practices is deemed unjust. In language tests, Shohamy states that there is a discrepancy between the test construct and the real-life construct of what constitutes language knowledge. While I recognise that the artificial nature of assessment has limitations on how close to real-life interactions assessments ultimately can be, there is a case to make that if assessments are to exist that these should best reflect the sociolinguistic reality in a way which is inclusive to as many people as possible. The use of a construct which shows language as stable and fixed and uses a standard native variety as the benchmark unjustly disadvantages ELF speakers (see Section 2.11 on Ethical considerations).

The interaction during the AOP assessments and classroom observations which feature in this study are ELF interactions. The communication, as with much communication in UK Higher Education, is in English between speakers of different first languages and often with diverse multilingual repertoires; Jenkins and Leung (2019) call international higher education 'a site of ELF communication *par excellence' (p.12;* italics in original*)*. There has been an acknowledgement by some language testers such as Tim McNamara (2014) and Linda Taylor (2014) that ELF research challenges the L1/L2 paradigm, but there is still much work to be done to consider how language assessment may be influenced by ELF research. Scholars in language testing have long questioned the use of native speaker performance on tests as a means of validating test items as this practice assumes uniformity of native speaker performance (e.g. Alderson, 1980; Bachman, 1990). Despite long-standing dialogue on the use of an L1/L2 distinction, McNamara (2014) calls for 'a radical conceptualization' due to slow progress in redefining the benchmark against which assesees' performances are judged. Successful communication can no longer be judged as native speaker competence:

> The distinction between native and non-native speaker competence, which lies at the heart of the [communicative English language testing] movement, can no longer be sustained; we need a radical reconceptualization of the construct of successful communication that does not depend on this distinction. (p.231)

If native speaker as the target is the point of contention, a compelling case has been made to discredit the practice of using an idealized native English speaker as the model. The myth of the native speaker as a relevant or desirable goal has been dispelled by research showing that communication breakdowns can be attributed to L1 and L2 users (Kim and Elder's 2009 study looking at aviation communication); L1 speakers' varied performance (including at the lower range) on gateway assessments from which they are exempt – especially performance-based assessments (Hamilton et al, 1993) and work in ELF showing successful communication which does not refer to a 'standard native variety' (notably Jenkins et al., 2011; Jenkins, 2000; Seidlhofer, 2011). There is also the persuasive argument that factors other than language use such as motivation and stage confidence factor into successful communication and academic performance (as discussed in 2.8.3 above). For instance, both L1 and L2 users of English report the need to make adjustments in spoken communication and difficulties in doing so in ELF interactions in Anglophone EMI settings (e.g. Margić, 2017; Dippold et. al, 2019). However, the case for a complete dismantling of a L1/L2 distinction is weakened by research, including in the field of ELF, which reports on apparent differences in strategies used by L1 and L2 users (e.g. Margić, 2017). More research is needed which investigates the credibility and usefulness of such distinctions being upheld. Shared linguacultural knowledge offers an awareness of local conventions which increases the likelihood of successful communication (Widdowson, 2020), however caution must be exercised not to overstate the shared linguacultural knowledge in particular groups of people.

Different approaches to the question of how the phenomena of ELF should shape future assessment practices have been, primarily conceptually, explored, and are in early stages of conception. To solve the lack of construct validity in current language tests using a stable standard variety as the language construct, Shohamy calls for ELF to be used as a central construct in language tests (p.587). Creating a construct based on skills deemed necessary in ELF communication has been tentatively explored (e.g. McNamara and Harding, 2018). Alternatively, the removal of language from a test construct has also been proposed by Hall's (2014) *Englishing* approach which calls for not focusing on language but what people can do with it. Jenkins and Leung (2019) recommend that all large-scale standardized tests which act as gatekeeping mechanisms to university study should be eradicated. Instead, local, discipline-specific assessment tasks should be conducted at time of admission. These tests would involve input in the design from disciplinary tutors and test developer consultants. Students would be given self-assessment tools and would themselves decide their readiness for academic study through the medium of English. The core criterion against which readiness would be judged is 'successful communication in situ', according to Jenkins and Leung. This would in itself involve a pre-defined

notion of successful communication and decisions would need to be made on how local the assessment would be (discipline level, programme level or more local).

In this research, I demonstrate that there is a strong case for student input in assessment decision making processes. However, I strongly believe that teacher assessors' involvement can be a force for good, not only in the assessment design but also in the assessment outcome. Placing the admission decision solely in the hands of people already willing to pay fees would I fear not do much to redress power imbalances at work. Leaving external pressures aside (e.g. familial pressure; employment prospects), as humans, we may not have the ability to determine our readiness even with a degree of knowledge as to specified tasks. Self-assessments can be flawed for those of different competence levels depending on task difficulty (Dunning and Kruger, 2000; Burson, Larrick and Klayman, 2006). Disciplinary tutors, with, in most cases prolonged exposure to various practices in academia, are likely to possess deeper understanding of the qualities which facilitate students' success in the host academic community, and are a vital resource in reaching the fairest decision feasible. However, disciplinary tutors may lack expertise regarding the communicative characteristics and demands of tertiary education (Hyland 2006; Airey 2012). The role of EAP practitioners and contemporary EAP practice deserves close attention.

 EAP practitioners, many of whom have a knowledge base related to discursive and communicative practices in Higher Education, can be a valuable stakeholder in the process. Rather than reinventing the wheel, we may find the seeds of alternative practice are already sprouting: in-house EAP assessments may not provide solutions, but I argue that they can offer much to the debate on alternatives to large-scale standardized assessments. The gaps in documentation and our understanding of what is assessed in EAP assessment is an area which requires close attention. These lacunae and their significance are explicated in the following section.

2.9.3   The Construct in EAP Assessment

There is a wealth of EAP assessment practice, namely in-house tests, however little EAP assessment theory exists that has not been developed by language testers (Schmitt and Hamp-Lyons, 2015). Schmitt and Hamp-Lyons have called for more work in assessment by EAP practitioners to help solve the issue of 'the construct of EAP in EAP assessment being under-defined and under-theorized' (p.3).

EAP practitioners have conducted research on AOP practices in the target domain (e.g. Levrai and Bolster, 2015), but little work has been done, or published at least, on what is happening on EAP courses. While there has been research conducted on AOPs in EAP settings most of this has

focused on AOP instruction rather than assessment tasks. Research conducted on AOPs on EAP courses has included the effectiveness of EAP instruction on skills necessary for AOP delivery (Bankowski, 2010) and teaching of the language for AOPs (Boyle, 1996; Zareva, 2009). Studies conducted on EAP AOP assessment practices is less common. An exception is Januin and Stephen (2015) used textual (documents: tutor review form and course textbook) and ethnographic methods (interviews with four teachers and 15 teachers and observational data from 32 video recorded presentations) to unearth the discourse competence elements ESL students are required to employ when delivering an AOP based on a written essay on an EAP course in Malaysia. The findings indicate that there are three types of discourse competence involved in delivering the AOP: public speaking, use of the oral presentation structure, and linguistic knowledge. Januin and Stephen did not investigate substantive processes involved in AOP gateway assessments, however. Few studies have investigated assessor and test taker decision-making processes in high-stakes classroom-based assessments (Hill and McNamara, 2011). Hill and McNamara (2011) and Rea-Dickins' (2001) work on assessment in secondary contexts with their focus on processes and adoption of a grounded approach has informed the focus and approach taken in this study in tertiary contexts (see Chapter 3 on methodological approach). I consider EAP programmes with gateway assessments as crucial sites to examine.

One way of framing academic performance and the construct in EAP is through the notion of 'ability to cope' (Fulcher and Davidson, 2007). As was seen in the opening quotation in Chapter 1 from Tracey, this is a driving concept in the teacher assessor decision making in S2 field sites. This term connotes the ability to deal with diverse situations which may be encountered in the target domain. On general EAP courses, EAP practitioners are often tasked with developing skills which may be applied in a variety of tasks in diverse settings. What constitutes 'readiness' or 'ability to cope' has been described as involving more than language proficiency (Schmitt, 2015). Gilliver-Brown and Johnson, 2009; cited in Wilson 2016) found that factors such as academic background, cultural expectations and prior cognitive skills development, as well as language proficiency, were factors contributing to a student's ability to cope with the demands of tertiary education (as discussed in Section 2.9.2 above). Furthermore, L1 speakers of English encounter challenges in coping with university study (Dippold, 2019), demonstrating that general language proficiency may not be a decisive factor in academic success. Furthermore, Academic Literacies scholars highlight the importance of considering the influence of epistemological stances and identities on practices, which vary from one setting and person to another (Lea and Street, 1998, 2006). Having the language skills to communicate is one aspect of successful performance of academic tasks and this is reflected in developments of notions of EAP (as discussed in Section 2.4).

Developing knowledge of what is happening in AOP assessments in particular contexts develops genre knowledge, which can improve teaching quality and curriculum and materials design, as well as assessment design and implementation, but it also raises core existential questions for the field of EAP. What happens in EAP assessment practice speaks volumes as to what counts in EAP and how it is conceptualised. The importance of understanding assessment is crucial as 'the spirit and style of student assessment defines the *de facto* curriculum' (Rowntree, 2006, p.1; italics in original). In a study (Elder, 2016) which investigated what clinical supervisors value in English for health communications, there were recommendations made for changes to be made to the criteria used in the Occupational English Test. Examining what EAP practitioners value in gateway assessments provides valuable grounded theorisations as to EAP constructs, which reflect current practice and resonate with practitioners.

There is a potential that practices of EAP practitioners is not being documented in research, especially as the EAP practitioner often 'operates on the edge of academia' (Ding and Bruce, 2017). Therefore, the field of EAP may not be sufficiently or accurately recognised or depicted. I believe that from conducting and publishing research on assessment practices, EAP practitioners' and students' experiences and the complex construct of EAP can be rendered more visible to each other, to those in academia and to wider society. Rendering practices visible affords greater accountability and reach of EAP practice. For instance, EAP practice may offer profound insights into alternatives to large-scale direct entry tests hence contributing to contemporary debates in critical language assessment.

Entangled with the notion of a construct are the concepts of validity and measurement, concepts which undergo a critical appraisal in the next section.

2.10   Validity and Measurement

Validity theory is pivotal to psychological, educational and language testing. A great deal of debate ensues regarding whether particular constructs or attributes are observable, theoretical or conceptual (Newton and Shaw, 2014). Measurement has been the central focus for validation for many validation theorists, however the relevance and validity of the measurement paradigm has been contested.

2.10.1   Rethinking the Measurement Paradigm

In early validity theory, there was a tripartite model of validity which encompassed criterion-oriented validity, content validity and construct validity. Samuel Messick in 1989 proposed a

unified framework of validity in which different types of evidence contribute to construct validity. In doing so, Messick changed the fundamental understanding of validity. A modified, and oft quoted, definition of validity was proposed:

> an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (1989, p.13).

Messick's definition is useful as it centres on the interpretations not the properties of a test and covers value implications, social consequences, relevance and utility (Chapelle, 2012, p.24). Validity is the ability to justify inferences made based on a test score. For instance, does a cut off score of 50 (pass mark) in an EAP AOP assessment indicate the test taker can participate on their prospective courses?

A quest to simplify the validation process was undertaken after Messick's unitary approach. Kane (1992; 2013) makes the case for an argument-based approach to validity. The view of validity as interpretation has gained widespread acceptance among validation theorists, however to date there has been little practical implementation of Kane's approach (Newton and Shaw, 2014, pp.140-5). Kane's approach emphasizes that validation is a process of constructing the best argument with the evidence available and being open to shifting the argument in light of new evidence or alternative explanations. Kane (1992) posits that 'any test-score interpretation always involves an interpretive argument' (p. 527) and the best assessors can do is present an interpretive argument as highly plausible. An interpretive approach has flexibility as it does not necessarily entail association with formal theories. However, Kane's approach still claims that validity is a 'property of a measurement procedure' (Newton and Shaw, 2014, p.158).

A number of contemporary validity theorists have taken issue with the notion that attributes in assessment are in fact *measurable*. The notion of measurement itself must be interrogated, they argue. With the presumption of measurability, without empirical proof, the concept of validity is itself invalid (Michell, 2009). Recent theories on concepts such as performance, competence and communication (discussed in Sections 2.8.3-2.8.5) pose serious questions as to the ability to predefine and measure abilities which may emerge in communication.

Researchers in classroom-based assessment have grappled with the question of how conceptions of validity and the measurement paradigm often applied to large-scale testing may and may not apply to less standardized assessment. The approach to validity which is espoused is relative to the context in which stakeholders operate and purposes of assessment. This research explores educational assessments based in classrooms, therefore notions of validity in such settings are discussed next.

2.10.2 Validity in CBA

Classroom-based assessment is defined as 'non-standardized local assessment carried out by teachers in the classroom' (Leung, 2005, p.871). Classroom-based assessment covers a wide range of assessment procedures. In the early 2000s, Tim McNamara called for a paradigm shift in assessment research so that research was made more applicable to learners and teachers. He proposed an expansion of the notion of assessment to include 'any deliberate, sustained and explicit reflection by teachers (and by learners) on the qualities of the learner's work' (McNamara, 2001, p.343) and that track development over time. CBA may constitute formal or informal assessment opportunities (Rea-Dickins, 2001). There may be formal assessment episodes such as a designated assessed task, but also informal assessments (observations made in class) which gather information about student capacities which may or may not be explicitly labelled as assessment events.

Assessment information can be gathered for a range of, often competing, purposes. Bonner (2013) points out that classroom assessments often have a learning purpose as well as a measurement purpose. Bonner argues that aspects of traditional validity theory may be applicable if measurement of learning is taking place, however other aspects of traditional theory (such as decontextualisation, standardization and pseudo-objectivity) will not be relevant for learning purposes and due to the characteristics of the educational settings in which the assessments take place. The role of context, the focus on learning, the longitudinal and diverse assessment windows and the virtues of subjectivity in CBA are factors cited which indicate that the traditional psychometric approach is not considered a relevant general heuristic for designing and implementing assessment practices in some settings (Moss, 2003; Brookhart, 2003).

The stark difference between large-scale tests and classroom assessment is the role of context. Traditionally, context is considered as part of construct-irrelevant variance (Fulcher and Davidson, 2007, p.25). Pamela Moss (2003) has cogently argued that the classroom context is part of the construct. Moss (2003) views learning from a sociocultural perspective in which 'learning is perceived through the changing relationships among the learner, the other human participants, and the tools (material and symbolic) available in a given context' (p.14).This context, as she describes it is a social one and a learning environment. Teachers build a picture of the learner by how they interact in this environment. Sundrarajun and Kiely (2009), in their investigation of collaborative talk in classrooms linked to AOP tasks, found that AOP tasks have the ability to create meaningful learning opportunities. The AOP performance assessments featured in this study also have a strong formative dimension (see analysis Sections 6.7.3.4 and 7.8 and discussion in Section 8.6.1).

Socio-cultural phenomena it has been argued become part of the test construct in educational assessment and CBA. Wolfersberger (2013) conducted an ethnographic study of four students tasked with completing an integrated writing-from-readings classroom-based assessment. The ethnographic data revealed that the students' task representations varied and contributed to a successful or unsuccessful test outcome. Wolfersberger argues that task representation, a mental model of the final product, should be included as a component of the test construct in multimodal dynamic classroom assessment. Meanwhile, Moss (2003) argues that the unit of analysis in classroom assessment, instead of the individual is the social situation. 'The emphasis on individual scores masks the complex role of the social context in shaping those scores and the interpretations they entail' (p.17). The EAP classroom interaction has the potential to provide assessment opportunities which are used to make inferences as to how students may cope with potentially similar interactions and tasks with future peers and teachers in the target domain. An analysis of rating discussions (see Chapter 6) shows that informal assessment opportunities enter EAP practitioners' score-reaching talk in S2 field sites.

Advocates of teacher assessments argue that large-scale, external, standardized tests are dehumanizing and demoralizing for teachers and learners and have an adverse effect on learning (Van Lier, 2004, p.29). Johnston (1989) criticizes the quest for objectivity in assessment decisions and describes it as harmful as 'it hides the political nature of the activity in a veil of science' (p.512). Johnston argued that strategies suggested by Lincoln and Guba (1985) in the context of naturalistic research be used to strengthen the plausibility of teachers' interpretations. These include: prolonged engagement in the classroom context, providing different sources of evidence, seeking reaction from colleagues, and leaving an audit trail (these strategies are discussed in relation to S2 findings in Section 8.6).

The situated, dynamic, emergent nature of local assessment calls for alternative perspectives on validity which acknowledge the prominent social dimension of CBA. Questions remain as to whether the concept of validity can be applied to the complex assessment interactions and decision making. Chapelle (2012) encapsulates the hybridity of conceptions of validity:

> In a world where audiences are characterized by hybrid influences of culture, academic traditions, and disciplinary affiliations, the idea of a single conception of validity may be as much of an abstraction as the conception of the native speaker. (Chapelle, 2012, p.31)

Validity is conceptualised in various ways and relative to the context. Using the paradigm of large-scale testing in educational, classroom assessment, while applicable in some cases, may be harmful in others.

Validity in assessment circles has been extended to the domain of ethics (e.g. Messick 1979).

I see validity and ethics as enmeshed together and now consider issues of ethics, fairness and justice in assessment.

2.11 Ethics, Fairness and Justice

Foucault (1975, p.184) posits that examinations are mechanisms which facilitate disciplinary power and control. Tests put one person or persons, and their particular values, in a position of authority in making a judgement with potentially huge social consequences for the examinee. Admittedly, the harm they do may well be greater than the good for given individuals, especially when particular ontological, epistemological and ideological approaches are uncritically applied. I share Liz Hamp-Lyons view that all testing is a political act (Hamp-Lyons, 1997). I believe that assessments ultimately put particular values in a position of power, but that tests have the power to do good as well as harm. Elena Shohamy (2001), while critical of many tests, powerfully points out that there are instances when tests improve opportunities for access to education. Assessment has the ability to create a fairer and more democratic process than already exists, regrettably not universally fair however.

Tests may challenge or bolster the knowledge valued by society's elite and maintain or subvert the status quo. Is the status quo ethical? What is ethical practice? Who decides what ethical practice is? Fulcher and Davidson (2007) view ethical practice as 'the constant exercise of self-questioning and open debate' (p.140) and the 'extent to which our work contributes to the opportunities which our tests open up to all citizens of our society' (p.142). They argue for an effects-driven approach which values dialogue with stakeholders. As discussed in the previous section, the quest for objectivity and viewing validity as a scientific endeavour, hides the political nature of assessment. Stakeholders on the contrary, by discussing the social and subjective in decision making, have scope to consider ethical implications. In FS1 and FS2, the individual assessors when making assessment decisions consider how fair their decision making is in critical episodes (see Section 5.7 below). Investment in the decision making, opportunities for dialogue and working conditions conducive to deep thinking is crucial to ethical practice (see Sections 5.7, 8.6 and 9.2.3).

The current admission policies of UK universities raises ethical questions. Who and what to assess in gatekeeping tests is fraught with complexity and quandaries. While support aimed at meeting communicative demands required in UK HE is jointly offered to L1 and L2 speakers of English in an increasing number of UK tertiary institutions, gatekeeping assessment practices continue, on the whole, to separate L1 and L2 speakers of English dividing students down linguistic and cultural lines. It seems this practice, to some degree, infers that home students are a homogenous group

with a shared typical standardized language repertoire, cultural and educational background. It also indicates that language proficiency is deemed a key driver in successful academic performance. While a base of linguistic skills is considered crucial for successful academic performance, an emphasis on the linguistic (especially at particular levels of proficiency, domains and tasks) has been challenged (e.g. Graham 1987, Canagarajah, 2018). The justification for in many cases granting entry via different routes for L1 and L2 speakers needs to be investigated to ensure equitable access for all students. EAP assessments are one route to entry and require research into practices in order to contribute to the debate on reasonable expectations and parity in entry requirements for the entire prospective student body. Developing a closer understanding of the ethical dimension of testing involves investigating the essence of what is happening in assessment practices. An ethnographically-oriented investigation of key stakeholder practices has the power to unearth the factors which shape individual assessment practices. An investigation of current micro practices can contribute to this long-standing debate as to what fair and just practice may involve in gatekeeping decisions.

Universities in the UK operate in a neoliberal system which sees universities as ranked competitive entrepreneurial organizations offering a service. This manifests as students becoming paying clients and consumers and staff performance managed and measured. Teacher assessor and student assessee micro practices are shaped in ways by such larger forces at work. Assessment practices are influenced, and potentially muddied and tainted, by political and socioeconomic factors. EAP practitioners are clearly subject to the forces of marketization and Hadley's (2015) critical grounded theory of EAP units indicates that the position of EAP practitioners forecasts the plight of disciplinary tutors – EAP units act as a barometer of the extent to which a university has been neoliberalized (2014, p.164). Ding (2016) puts forth the argument that EAP is a *product* of neoliberalism with its focus on attracting international students and as a money-making machine for universities. A great number of EAP practitioners confront and experience an intense moral conflict in their practice: they believe in supporting students in their studies but feel deeply uneasy about dividing students along linguistic lines and 'the overt use of the international student 'market' by governments to shore up the finances of universities is an embarrassment' (Hamp-Lyons, p.A2).

EAP practitioners Tracey and Georgina in field site two have a fascinating discussion which addresses macro issues. The discussion these teachers have (see Section 5.7) sheds light on a culture of pressure to pass students in their institution resulting it seems from the effects of neoliberal ideology on universities. The practices theorized in this thesis demonstrate what counts to teacher assessors and how they justify the ethicality of their decision making under structural constraints. Admittedly, a small part of the data sets directly reflect the macro factors raised in

this section but these are telling episodes indicating influential features of the context, which shape how stakeholders approach their practice at classroom level.

In this study, I use McNamara and Ryan's (2011) distinction between fairness and justice in assessment. McNamara and Ryan (2011) use the term *fairness* to relate to issues internal to the test and *justice* for aspects external to the test. The distinction between the two concepts is clarified in the following quotation:

> …even a test that was as fair as it could be made to be, carefully designed, validated, its areas of weakness identified and remedied as far as was humanly possible, might still involve issues of justice in the very fact that it was used at all, for example, as serving a misguided or discriminatory policy.
>
> (McNamara, Knoch and Fan, 2018, p.19)

In other words, if the test construct reveals privilege to native speakers over non-native speakers this would be an issue of *justice*. An issue of construct-irrelevant variance, such as test-takers being assessed on different abilities, would be classed as unfairness, according to McNamara and Ryan's distinction. This thesis focuses predominantly on fairness, but justice implications of findings in this study are considered in section 9.4.

2.12   Chapter Summary

The bricolage presented in this chapter has covered a number of important research streams and concepts which both underpin and have explanatory power in this research. My exposure to EAP practice through data analysis required a swift and evolving construction of a novel theoretical framework which constitutes interrelated diverse theories on oracy, performance assessment, EAP, genre analysis, multimodal communication, assessment constructs, validity and fairness.

In Sections 2.2-2.3, I reviewed scholarship undertaken related to oracy demands in HE, summarising a growing body of qualitative research in a hitherto marginalised area. I also shared the inclusive (but not exhaustive) notion of oracy which has evolved during the research process. I, then, in Sections 2.4-2.6, reviewed research in the field of EAP and Genre Analysis, unpacking the concepts of EAP and genre. In Sections, 2.7-2.9, I devoted attention to theories in second language assessment. In Section 2.8, I summarised key trends in language assessment. Following this, I focused on developments in thinking regarding the role and nature of the construct in performance assessment drawing on a range of concepts (e.g. poststructuralism, performativity, transmodalities) to make the case that communication is emergent, dynamic and difficult to encapsulate in a pre-fixed construct. The review then turned to pivotal concepts in assessment – validity and measurement (Section 2.10). I aligned myself with Moss' vision of CBA which creates its own notions of validity potentially removed from large-scale standardized testing and Johnston's view that objectivity has great potential to harm by not acknowledging the social

dimension of decision making in CBA. I do not preclude that the measurement paradigm and features of large-scale tests may retain relevance in some classroom-based assessments, however. The final section of this chapter raised the ethical dilemmas confronted in the act of language+ assessment. Admittedly, I have assembled a diverse and complex conceptual and theoretical framework. Upon reading the remaining parts of this thesis, I hope this complex enmeshed web will become increasingly instructive.

All of these areas have profound standing in this research and pervade most aspects of the project, but due to the linear nature of text, they will be brought to the fore at different points in this thesis. The next chapter provides a detailed discussion of the study's design and the rationale behind the methods used in order to become more informed as to the processes involved in AOP assessment practices.

Chapter 3   Research Design and Methodology

3.1   Introduction

This chapter details the research design and methodological decisions taken in this study. I make the case for a qualitative, grounded and ethnographically-oriented research approach to investigating AOP assessment practices. I set out the argument that such an approach is compatible with the object of enquiry: the under-researched area of the processes engaged in by teachers and students in AOP assessment in tertiary education contexts.

In writing this chapter, I have grappled with identifying and making concrete what I have done with a clear sense of design and elements which emerged in the course of action. I endeavoured to follow leads which captured data on important issues in understanding teacher and student approaches to AOP assessments. In this chapter, I first discuss the epistemological and ontological discourses related to the established research base on assessment practice and detail the philosophical underpinnings of this study. In Section 3.3, an overview of the study is provided. In the latter part of Section 3.3 and in Sections 3.4-3.5, I account for the blend of CGT and ethnography. I provide a description of and rationale behind the data collection and analysis methods I adopted, as well as details of the S2 fieldwork sites and participants in Sections 3.7-3.9. The chapter ends with a focused discussion of validity and ethical considerations. The whole chapter, however addresses questions related to the quality of the research, meaning the validity of the research design and methodology is in fact discussed throughout the chapter.

3.2   Issues of Epistemology and Ontology

Epistemological developments in language education have triggered expansion and shifts in methodological approaches taken to language teaching and assessment research and practice. With a distinct uptake of Sociocultural Theory (Vygotsky, 1978) and its emphasis on context by educators and researchers in language education, there has been a notable swing from a positivist philosophical position towards naturalistic, interpretive approaches in understanding language learning, teaching and assessment practice (Moss, 1996; Wajda, 2011; Tsushima, 2015). Since the 1990s, the importance of context, culture and values in shaping assessment practice has been highlighted (e.g. McNamara, 2001), which has sparked epistemic and ontological shifts in perspectives taken on language assessment practice. As discussed in the previous chapter,

73

classroom dynamics add to the complexity of assessment practices due to the heightened social dimension of practice in a teaching environment with often prolonged contact between students and teachers (Fulcher and Davidson, 2007; Green, 2014) (see Section 2.10.2). Most notably, there has been an increase in qualitative methods being used in the study of classroom-based language education and language assessments, either in mixed-method approaches or (almost) exclusively qualitative studies (Tsushima, 2015).

I position my research as an educational research project (Bassey, 1996). I situate this study within the paradigm of constructivism which moves beyond interpretivism by stating that the researcher is also part of this social reality and therefore part of its construction. The main tenets of constructivism are that reality is a social construct (Robson, 2002) which entails that it is subjective, evolving, emergent and under construction. This view is in contrast to the positivist stance which purports that there are stable, defined truths to be discovered by research (Robson, 2002).

There are two main reasons for conducting a grounded, qualitative study within a constructivist paradigm in this research. First, the lack of research in EAP assessment (as discussed in Section 2.3 and 2.9.3) entailed a grounded approach. The use of a naturalistic (Lincoln and Guba, 1985), rather than rationalistic mode of enquiry, has been chosen in this study as it affords the opportunity to explore this undertheorized area of practice which is highly contingent on understanding the actions of people *in situ*. While the boundaries between qualitative and quantitative research are mobile, qualitative research is described as involving an interpretive and naturalistic approach (Denzin and Lincoln, 2003, p.3). Second, the phenomena under investigation has been shown to be a complex phenomenon involving numerous unknown variables, factors or phenomena at the outset. Qualitative research, in comparison to quantitative, is intent 'on expanding rather than controlling variables' (Holliday, 2007, p.29). Qualitative data was gathered because this study seeks to understand AOP assessment practices and constructs from the perspectives of teachers and students, allowing for the myriad of unmeasurable nuances involved in the phenomenon.

Researchers investigating decision-making processes and strategies in classroom assessments in previous research have opted for qualitative research designs. Rea-Dickins'(2001) study of teacher decision making in primary school classroom assessment practices for learners with English as an additional language took a qualitative, grounded approach within an interpretive paradigm. The report cited data from a larger project which used classroom observation, recordings of classes and interviews with teachers and language support staff to develop closer understanding of processes at work in classroom assessment. The grounded qualitative approach

resulted in a rich account of classroom assessment practice as a complex undertaking which is shaped by learning, teaching and administrative functions. It also provided a working framework for analysing a range of strategies in classroom assessment. Hill and McNamara's (2011) study similarly focused on decision making in classroom assessment. It tracked two classes of young learners studying Indonesian as a foreign language. Akin to Rea-Dickins, Hill and McNamara adopted a grounded, qualitative approach and used ethnographic principles and methods (participant information and case studies). Hill and McNamara's study yielded diverse data sets which encompassed a wide range of documents (e.g. rubrics, teacher notes, role-play scripts) and task types (including oral presentations) for analysis. This resulted in a comprehensive framework which encompassed the multifaceted areas and issues involved in CBA research and practice. This current research departs from Rea-Dickins' (2001) and Hill and McNamara (2011) in that it focuses on documenting processes specific to gateway decision making in a specific genre (AOPs) on EAP courses in UK tertiary contexts.

In this research, I reflect individual constructed realities using rich interpretive data gathered in qualitative surveys (in S1) and from fieldwork in natural settings (in S2). The features of qualitative inquiry are relevant to my study and embraced, but are also put into practice with some caution. In order to provide a sufficiently convincing account of events which have a degree of practical use in the world, the art lies in acknowledging that the qualitative researcher offers an *interpretation* of reality whose legitimacy is understood in terms of its plausibility and based upon the transparency of the account of the empirical data obtained (Bogdan and Biklen, 2007, p.26). The quality of the research, described by the term validity in this research, is addressed throughout the remaining sections, culminating in a focused discussion in Section 3.10.

The next section provides an overview of the study's research design.

3.3   Research Design Overview

This part of the thesis contains an overview of the research design (expanding on the overview provided in Section 1.4). The research from the outset was guided by a broad research question: What is happening in EAP AOP assessments? In order to refine the focus of the study a qualitative survey was conducted before engaging in fieldwork at two sites. Qualitative surveys were conducted to gauge the extent to which AOPs are used in UK HE practice and what is entailed in producing AOP assessments on subject modules and EAP modules. The Survey Study indicated that AOPs on subject courses and in EAP settings warrant investigation (see Chapter 4 for analysis of S1 data). The fieldwork study informed by ethnographic principles builds on The Survey Study

to explore how AOP assessment is done in two EAP settings on an IFP and Pre-Master's EAP modules (see Chapters 5-8 for analysis of S2 data).

The overall guiding research questions which have progressively emerged, from adopting an ethnographic orientation to the research, through the course of the study are provided below:

Research questions

1. In what ways are AOP assessments used in UK Higher Education?
2. What is the EAP AOP assessment construct?
3. What factors determine the EAP AOP assessment construct?

It has proven difficult to reduce the richness of the phenomena I am researching into contained and a limited number of precise questions. According to Holliday (2007, p.28), this is to be expected as qualitative research expands rather than controls variables. Throughout the research, the central research questions changed significantly. I also decided to devise a series of specific sub-questions for each of the studies (S1 and S2) in the project. The main and sub questions evolved during phases of the project; those questions in Table 3 below are the final revised versions which acted as loose guides to data collection and analysis. The study stages, overall purposes, central research questions, sub-research questions which guided the studies and methods used to obtain data, are summarised below in Table 3:

Table 3: Detailed Research Overview

A) Study name	Study one (S1) The Survey Study	Study two (S2) The Fieldwork Study
B) Description	Qualitative surveys documenting AOP formats and stated practices on subject programmes and EAP programmes at UK universities.	Qualitative fieldwork investigating AOP assessment practices on EAP modules at two sites.
C) Purpose	S1 gathered information on factors which shape AOP purposes and communicative demands in UK HE. This is used to inform the focus of fieldwork in S2.	S2 obtained and analysed data on EAP AOP practices In doing so, the research provides accounts of teacher and student decision making in AOP assessment tasks.
D) Main research questions	1. In what ways are AOP assessments used in UK Higher Education? 2. What is the EAP AOP assessment construct? 3. What factors determine the EAP AOP assessment construct?	
E) Sub-research questions	S1 subject module data RQs: 1) What are the purposes of AOP assessments on subject modules?	Guiding research questions S2: 1) What are the teacher and student conceptions of the EAP AOP assessment purposes on their course?

	2) What criteria are used in subject module AOPs?
	3) What language use is viewed as appropriate in subject module AOP assessments?
	4) What challenges do students and teachers face in AOP assessments on subject modules?
	
	S1 EAP module data RQs:
	
	   1) What are the purposes of EAP AOP assessment?
	   2) What criteria are included in rating scales designed for use in EAP AOP assessments?
	   3) What emphasis is placed on linguistic and non-linguistic factors in EAP AOP assessments?
	   4) How is language use treated in EAP AOP assessments?
		2) What do teacher assessors do in EAP AOP assessment practice on their course?
		
		3) What factors shape teacher assessor AOP assessment practice on their course?
		
		4) What factors shape student perspectives on EAP AOP assessments on their course?	

F) Methods	1. Website page survey of AOP assessment tasks 2. Analysis of AOP task brief documents 3. Analysis of AOP marking criteria 4. Questionnaire for subject lecturers 5. Questionnaire for EAP practitioners	1. Participant observation: lessons, rating discussions, AOP assessments 2. Field notes made throughout course of fieldwork 3. Video-recorded AOP assessments 4. Audio-recorded rating discussions between two teacher assessors 5. Audio-recorded interviews with main teachers and student informants 6. Document analysis: syllabus, marking criteria, lesson materials	

The research entailed investigating a specific phenomenon across settings with a focus on processes. This meant that the phenomena and research questions were suited to a grounded theory (hereafter GT) approach. Furthermore, as little work has been conducted into AOP assessment practices in HE (as discussed in Section 2.3), I turned to predominantly inductive approaches to data gathering and in the initial stages of analysis. CGT and ethnographic principles have shaped the course of the study as they fit with the paradigmatic and epistemic approach to the study of AOP assessment phenomena (see Section 3.2). CGT guidelines stipulate that the data is used to shape interpretations and that a close examination of data is encouraged before drawing upon any extant theories (Charmaz, 2014). Kathy Charmaz encourages constructivist

grounded theorists to learn about the world under study before entering in ways other than conducting an exhaustive literature review:

> Similar to a camera with many lenses, first you view a broad sweep of the landscape.
> Subsequently, you change your lens several times and shorten your focal points to bring key scenes
> closer and closer into view. (2014, p.27)

Kathy Charmaz's description captures the progressive focusing adopted in this research. I first conducted a 'broad sweep of the landscape' by conducting The Survey Study (see S1 in Table 3 above). The use of a range of data sets provided different angles on AOP practices and pointed to the 'key scenes' to be explored to better understand the processes at work in AOP assessment. The rating discussions between two assessors and interviews became the central data sets (See Table 3, section F, methods 4 and 5 in S2 column). These became the central data sets as they provided direct, rich and telling justifications behind assessment decisions taken by teacher assessors and students.

In a time when methodological eclecticism is increasingly embraced, researchers adapt and apply approaches and methods in unique ways to meet the requirements of the research task and skills of the researcher (Bamkin et al., 2016; Battersby, 1981; Pettigrew, 2000). Remaining true to paradigmatic heritage is of paramount concern and mixing of methods and methodologies is discouraged by some researchers (Hughes, 1990; Connell et al, 2001). Interplay between methods and methodologies is embraced by others (Miles and Huberman, 1994; Hadley, 2017). Hadley likens the research process to the interaction between celestial material and young stars in our universe:

> …methods act very much like the material found on the far reaches of a stellar system or
> cluster of protostars. Methods can be drawn into the service of other methodologies
> and research traditions. This eclecticism can be seen, for example, where coding
> methods often associated with grounded theory are used with other traditions such as
> case studies, action research, or ethnography (Heath et al. 2008). Therefore, academic
> discourse, methodologies, methods, thought experiments, and research investigations
> exist in systems of paradigmatic belief that are marked by continuous interaction.
> Symbolically, like the particulate matter in orbit around a protostar, new ideas,
> practices, and insights are constantly colliding and interacting with each other to form
> new knowledge (2017, p.17)

Particular research phenomena call for an open and flexible approach to grapple with their inherent complexity. I share Hadley's vision of research practice as an interactive, dynamic and constantly evolving process. I embrace the collision and interaction between methodologies and have fused CGT and ethnography in this project. The use of ethnography and GT to explore assessment practices has been advocated and implemented in more recent years (e.g. Hill, 2009),

but it is still a relatively alternative method in assessment circles, and notably in language testing ones. Rea-Dickins (2001) and McNamara and Hill (2011) use a grounded approach and ethnographic tools (discussed in Section 3.2 above). Hill's (2009) research uses ethnographic data gathering techniques, GT and Foucauldian tools and notions to explore the choices made by teachers in their assessment practice in New Zealand primary schools. Hill convincingly demonstrates that blending approaches to data collection and interpretation is useful in gaining fresh perspectives on classroom assessment practice. Paradigm proliferation (Lather, 2006) is a fundamental component of new knowledge creation.

From viewing AOP events through a grounded ethnographic lens we may gain new insights into how to think about and approach EAP AOP assessments. Charmaz and Mitchell (2001) advocate the use of GT in ethnography and vice versa. In the following sections, I describe in detail how I have applied characteristics of grounded theory and ethnography to the study of AOP assessment practice.

3.4   Application of CGT

Since 1967, GT (Grounded Theory) has become widely used (Timonen et al, 2018) with, according to bibliometrical data, two thirds of social science research using the methodology in full or partial forms (Titscher et al., 2000). Despite the prevalence of GT methodology in social science, it has been sidelined in Applied Linguistics (Hadley, 2017). Gregory Hadley's (2017) recent book 'Grounded Theory in Applied Linguistics Research: A Practical Guide' has been an invaluable source of information and comfort. It has given much needed clarity on the evolutions in GT methodologies and the opportunities and challenges in using GT. In the course of conducting this research, I have encountered misconceptions as to what is involved in GT studies and what grounded theorists do, namely due to the diversity and flexibility of the work of grounded theorists not being acknowledged. The main criticisms posited at GT research are that grounded theorists overemphasize inductive reasoning and too readily dismiss existing theoretical literature. The crucial characteristic of GT is that theoretical concepts are only admitted when they are deemed relevant to the data: as mentioned in Chapter 1, the data filters which established theories have resonance. Furthermore, grounded theorists using more contemporary grounded theory methodologies (such as constructivist or critical GT methodologies) work with abductive reasoning and retroductive (Hadley, 2019).

As mentioned in the previous section, I draw on the inquiry traditions of GT due to the fact this study explores an under-researched area and seeks to understand the actions and processes involved in particular settings (AOP assessment processes on courses which determine readiness

for academic study). GT methodology is suited to studies which begin as open-ended and exploratory in nature and do not work with a hypothesis at the outset. GT developed from Symbolic Interactionism which is described as 'a dynamic theoretical perspective that views human actions as constructing self, situation and society' which offers concepts for interpreting social actions and realities (Charmaz, 2014, p.262). The Symbolic Interactionalist perspective stresses the importance of language and symbols in creating and communicating meaning and actions. While Symbolic interactionism is the foremost theoretical viewpoint associated with GT, there are different branches of GTM which may suit paradigmatic, theoretical and methodological perspective of the researcher (Hadley, 2017).

The diverse 'family' (Bryant and Charmaz, 2007, p.11) of GT methodologies has evolved over the years and here I discuss Classical, Glaserian, Straussian and Charmazian approaches. Barney Glaser's and Anselm Strauss' Landmark book The Discovery of Grounded Theory - published in 1967 - urged researchers to create new theories from the bottom up from empirical field data and by getting close to the practice. In the 1990s, Glaser and Strauss' explanations on the philosophical underpinnings of grounded theory and coding procedures revealed a disparity in how they envisioned the progression of GT (Hadley, 2017, p.37). Strauss advocated less rigorous coding aimed at understanding aspects of an object of enquiry, while Glaser introduced layers of coding aimed at fostering objective explanations of a phenomenon which exists ready to be discovered (Charmaz, 2014). Glaser's approach aligns with realist ontological and epistemological beliefs (Charmaz, 2000; Hadley, 2017). Conversely, Strauss began to move towards an interpretive perspective (e.g. Strauss and Corbin, 1998). This move towards the interpretative paradigm reflects a shift in qualitative inquiry over the past fifty years. A branch of contemporary GT gaining in popularity, and principles I draw on, is Constructivist Grounded Theory (Charmaz, 2014). CGT is premised on the notion that there are multiple worlds and multiple ways of knowing instead of one reality out there to be discovered and that the researcher and participants co-construct the reality under investigation. Realities are constructed by both the researcher and informants. A resulting theory generated by grounded theorists is regarded as an interpretation and directed at creating situated knowledges rather than universal abstract theories (Haraway, 1991; cited in Charmaz 2014).

GT studies may sharply differ due to the ontological and epistemological stances and capacities and dispositions of individual researchers in investigating their particular objects of enquiry (Hadley, 2017, p.60). There are, however, underlying commonalities between different GT methodologies even with the great variety of applications. Charmaz (2018) describes GT as inductive, abductive, comparative, interactive and iterative. Hadley identifies core methodological characteristics of GTMs outlined in Table 4 below:

Table 4: Core Methodological Characteristics of GT (Hadley, 2017, p.59)

Reflexivity	Grounded theorists are open to the informants, the flow of the academic community and to their own unique research talents, cultural background and philosophical research standpoint.
Coding	The data is fractured and reinterpreted so that the grounded theorist can begin to work with it in a meaningful way. Coding and data collection take place simultaneously.
Constant comparison	Codes and their interpretations are constantly reflected on in memos and compared with new data as it becomes available.
Theoretical sampling	The investigation follows the theoretical implications of the data as it has been interpreted by the grounded theorist.
Theoretical coding	The theoretical concepts are reconstructed and then linked into a framework that will be meaningful to a comprehending discourse community.
Grounded theory	The theory will be useful. It studies the concerns of a particular group of people and the ways in which they deal with these issues. The theory results in heightened awareness, understanding, ability to predict, and/or a greater sense of control, first for the group that was studied but then later potentially for people in other situations, but who are experiencing similar problems or concerns.

Through coding, constant comparison, theoretical sampling and theoretical coding, grounded theorists raise the description of phenomena to abstract categories and theoretical explanation. Theoretical sampling and formal theory development is viewed as a crucial component of a grounded theory study. Charmaz (2014), however, recognises both as legitimate grounded theory studies research which uses CGT as an analytic framework to build categories and concepts and that which builds a fully-fledged theory. Table 5 below outlines 9 key actions grounded theorists engage in when conducting a GT study, according to Charmaz.

Table 5: Actions of Grounded Theorists (Charmaz, 2014, p.15)

1. Conduct data collection and analysis simultaneously in an iterative process

2. Analyze actions and processes rather than themes and structure

3. Use comparative methods

4. Draw on data (e.g. narrative and descriptions) in service of developing new conceptual categories

5. Develop inductive abstract analytic categories through systematic data analysis

6. Emphasize theory construction rather than description or application of current theories

7. Engage in theoretical sampling

8. Search for variation in the studied categories or process

9. Pursue developing a category rather than covering a specific empirical topic

In this study, I have implemented the core characteristics of GT methodologies as outlined by Gregory Hadley (see Table 4 above). I have also engaged in strategies 1-9 in Table 5 above identified by Kathy Charmaz. Theoretical sampling and search for variation could have been engaged in, to a greater degree, if I had collected data from additional field sites. The processes and categories generated in this study may be developed further in future research.

CGT has been particularly instructive as an analytical framework having facilitated this enquiry by providing systematic, flexible guidelines for analysing data (see more discussion in Section 3.9 on analysis). Kathy Charmaz's (2014) publication 'Constructing Grounded Theory', her interviews with Graham Gibbs (Charmaz, 2015) and her workshop at Lancaster University (Charmaz, 2018) have guided me through the implementation of CGT principles and techniques in this study. I found that GT methodology alone did not wholly fit my research agenda, however, and so I turned to ethnographic research.

Charmaz and Mitchell's (2014) paper was pivotal in illuminating the possibilities afforded when blending GT with ethnography. The article was published in the Handbook of Ethnography and argued that GT can be incorporated into ethnography. However, it is also argued that GT studies can benefit from ethnography. This present study, although difficult to ascribe a concrete label, is a grounded theory study incorporating ethnographic elements, however CGT and ethnographic principles are very closely balanced and blended. In the next section, I provide justifications for having used ethnographic inquiry traditions in the study of AOP assessment practice. In doing so, similarities and differences between CGT and ethnography are highlighted.

3.5   Application of Ethnography

Ethnography is described as 'a mode of looking' and 'set of dispositions' rather than a methodology (Hammersley and Atkinson, 2019, p.238). Ethnographic work involves the following set of dispositions:

1. Not jumping to quick conclusions, even though the aim is eventually to reach some conclusion;
2. Paying detailed attention to appearances, while not taking them at face value;
3. Seeking to understand other people's views without treating what they say as obviously true or false;
4. Examining the circumstances in which people act, including much that they may not be aware of themselves, yet without losing sight of what they *do* attend to; and
5. Avoiding making evaluations of people, processes, organisations, institutions, etc. while yet being aware of the evaluations that participants and informants make, and investigating the sources and consequences of these.
   (Hammersley and Atkinson, 2019, p.238)

I have approached the research with what Hammersley and Atkinson (2019) describe as an ethnographic sensibility, which in fact coheres perfectly with the sensibilities promoted by Charmaz for constructivist grounded theorists (Charmaz 2014, 2015, 2018). The set of dispositions outlined above are how I endeavour to approach interactions in many life encounters, however the main difference is in the amount of evaluations I may make in real life compared to in my encounters with informants and documents in a research capacity. The principle of suspending judgement and bracketing out background assumptions and theoretical concepts aligns with the important place of reflexivity and predominantly inductive approach in CGT. Construction and subjectivity does not entail that the account is not accurate (Hammersley and Atkinson, 2019, p.16). All qualitative research is created by humans and therefore subject to the researcher's cultural background, but acknowledgement of such influences and the ability to suspend particular political beliefs in the act of documenting what other people are doing is vital. The crucial factor is that the prime aim of research is to produce knowledge, not political activism, otherwise the interpretation may be distorted by the researcher's values, including their political beliefs (Hammersley and Atkinson, 2019). The research produced may very well enter the political realm, and the researcher may need to be attuned to, and in some cases publicly confront, the potential for their work to be taken up, and in some cases distorted, by people for political motives. A key tenet in both CGT and ethnography is that researchers approach phenomena with a view to minimising any preconceived ideas as to what will emerge in the data.

The ethnographic principle of gaining immediate access to the phenomena being researched has greatly influenced this project. CGT and ethnography alike seek to understand phenomena through the lived experience of informants. However, a criticism levelled at much GT research is the practice of describing phenomena from a distance, or from the outside. For example, a theory of how teachers make assessment decisions could be conducted purely by interviewing teachers and not observing real-time teaching and assessment processes in a GT study. Ethnography prompts the researcher to enter the pivotal scenes when action takes place, in order to understand what informants experience, rather than how they speak about particular phenomena (Charmaz and Mitchell, 2001). In this study, I observe teachers and students in their classroom settings, in the AOP assessment events and in rating discussions. Direct observation is crucial as informants may not share aspects which they take for granted or which may be difficult for them to divulge, or even of behaviour of which they are unaware. The actions of individuals may not be consciously acknowledged, therefore ethnographic work is needed if such unconscious behaviour is to be accessed (Hammersley and Atkinson, 2019, p.239). Being present at the events and spending time with informants in formal and informal settings allowed me to build a richer, embodied experience of the events and pictures of people from drawing on all my senses. As a result, I drew on the experience of the events as well as the recorded data. This ethnographic dimension has strengthened the plausibility and credibility of the accounts of AOP practices (see Section 3.10 on validity).

The study of not only the actions of individuals but the individuals and the local cultures in which they operate also meant ethnographic tools were instructive in making sense of AOP practices. This research focuses on local, in-house, performance assessments which are characterised by elements which are not pre-designed and controlled. The diversity of constructions, resulting from different teams in different HEIs creating their own strategies for managing performance assessment, means local cultures - sometimes the perspective of one local person - are highly influential. These features point to the need for capturing the ethnographic and constructivist. The theoretical categories built using CGT methods are important but their significance and meaning are retained by being interwoven with contextual analysis of actions and events. The description of the context is thicker and richer when incorporating ethnographic encounters.

Participant observation is often characterised as a core feature of 'pure' ethnography or an ethnographically-oriented study. Blommaert and Dong (2010) argue that the ontology and epistemology of ethnography require complete immersion and active involvement, which requires time to achieve. Shah (2017) includes long duration as a core aspect of participant observation. However, for others, short-term participant observation has its merits and place in

multi-method ethnographic research (Brockman, 2011; Pink, 2013). Furthermore, the researcher may have varying levels of involvement in separate sites and different activities within sites. Within the current research climate, shorter ethnographic projects have become increasingly commonplace. Pink (2013), therefore, considers the merits of shorter timescales of fieldwork in view of the wider project ecology.

I have adopted strategies to maximise the opportunities presented in short-term fieldwork and participant observation in multiple sites. One such strategy was to focus my project around a central event – AOP assessments. Jeffery and Troman (2004) proposed three models of strategies used when there are shorter periods of fieldwork which attempt to involve the nature and principles of ethnographic research: compressed, selective-intermittent and recurrent. My study uses the recurrent strategy – that is, my project is structured around a temporal phrase. I structured my research around the AOP event and rating discussions. These were the central episodes and I focused on observation and conducted interviews prior to and after these events (see Section 3.8.2 on staging of methods). Another strategy was to have relatively focused observation by focussing on actions related most closely to the formal AOP assessment and identifying key informants early in the fieldwork. I also observed informants in various settings and at different stages in the process (site visits spanned 4 months). I did not observe the class over the full academic year, the life cycle of the course, however (see chapter 9 for limitations). In future research, I would opt to conduct longer and more immersive participant observation if this lends itself to the environment and research agenda.

Brockman (2011) argues that negotiating a researcher role in participant observation is important in obtaining meaningful data and this can more easily be achieved in relatively structured and formal environments, such as classroom settings. During my fieldwork, I found that while it may be easier to assimilate into more formal settings and find a role, this may exacerbate other challenges such as building rapport. Opportunities to interact in less structured settings forms a vital means by which to build rapport and gain fresh insights into what is happening at fieldwork sites. The fieldwork in S2 involved a complex mix of encounters in which I could place myself on numerous points on an active member and periphery member (Adler and Adler, 1998) continuum. In the rating discussions, at points, I was positioned as almost a complete observer, while in some lessons I was more closely positioned as a complete participant for large parts of certain activities (see further discussion in Section 3.8.1 and 3.8.3.2). I do not feel I was at any point 'a complete participant' (Creswell, 2013, p. 167) but I was 'an impartial onlooker' (VandeVen, 2007, p.269) for most of the rating discussions for reasons of fairness. I discovered that moments of clarity and salient insights were shared in ethnographic conversations after the set interviews and rating discussions, in the café or walking across campus to class.

Constructivist grounded theorists, instead of studying one setting and group, investigate processes involved in one phenomena in different settings (Charmaz, 2014). In this study I have investigated the shared and divergent processes in two field sites and included ethnographic details to further understand the phenomena. The blend of CGT and ethnography results in a theoretical depiction of AOP assessment practices using analytical categories and thick description. I build theory but do not claim that this is a representation of contexts other than those which feature in this research. I instead, hope, rather than promise, that the analytic categories and accounts presented may point to patterns of action in other areas of practice and therefore be instructive in navigating complex phenomena and undertakings, especially in educational assessment contexts. In analysing the data, a tension has arisen between emphasising the complex and contingent nature of phenomena and the emerging need to describe, interpret and theorize in a way which makes some sense of assessment practices. To navigate and develop closer understanding of assessment practices has involved compacting the complex phenomena into digestible, instructive analytic categories.

In this research, I first conducted surveys to unearth potential areas to explore in the second, main phase of the PhD, which involved ethnographically-oriented fieldwork. Both CGT and ethnography start with a broad spectrum of data and progressively narrow the focus (Charmaz, 2014; Hammersley and Atkinson, 2019). I now turn to a discussion of S1 procedures which served to conduct 'a broad sweep of the landscape' (Charmaz, 2014, p.27).

3.6   S1 Procedures and Methods

The Survey Study -- S1 -- entailed stepping into unchartered territory. I set out to gather information on AOP assessment tasks. This involved conducting a survey of webpages for subject programme AOP assessment task information, sourcing and analysing AOP task briefs and criteria on subject programmes and EAP programmes and analysing responses to self-designed questionnaires targeted at subject lecturers and EAP practitioners. The next sections provide an account of the access negotiated, the methods employed and the steps in analysis which were applied to the data.

3.6.1   S1 Access

S1 involved negotiating access to assessment documentation and obtaining responses to questionnaires at UK universities. I decided to approach one university to gain examples of the documentation -- AOP task briefs and marking criteria -- in order to aid the design of the questionnaires. Gaining documentation in S1 on subject courses, was a small part of the PhD

study as a whole, and I wished to minimise any further ethics applications and ethical concerns which may arise from negotiating access to assessment documentation in numerous institutions. Therefore, I limited the assessment criteria document analysis to one institution. When undertaking the task of obtaining marking criteria and feedback sheets, responses varied. Four departments provided documents; other departments wished to retain control of these data for reasons such as concern about copyright issues involved in sharing bespoke marking criteria.

The questionnaire for subject lecturers (henceforth SLQ) was sent to institutions through admin contacts at ESRC doctoral training centres at UK universities. I received 45 full responses to this questionnaire. The BALEAP (EAP practitioner association) mailing list was used to ask EAP practitioners for copies of their marking criteria and to send out the link to the online questionnaire for EAP practitioners (hereafter EAPQ). The BALEAP professional network has proven to be an indispensable tool for gaining information for the project. 14 EAP practitioners completed the EAP practitioner questionnaire. Access to staff at universities has required careful consideration and planning and has revealed diverse access requirements in both S1 and S2 (see Section 3.10 on ethical considerations).

3.6.2   S1 Methods

S1 used three ways of gathering information on AOP assessments: website surveys, document sourcing and qualitative questionnaires. In Section 3.6.2.1, I describe and justify the analysis of webpages and documents related to AOP assessment in UK HE. I then, in section 3.6.2.2, provide details of, and the rationale behind, the use of qualitative questionnaires.

3.6.2.1   Website Survey and Document Sourcing

The website survey of subject course AOPs was conducted on webpages at one Russell group UK university to build a picture of the diverse types of AOPs at one institution. The website survey entailed searching through individual module assessment pages to find oral presentation assessment information. This information often included details such as whether the presentation was delivered individually or in a group, the weighting and in some cases a brief description of the task (e.g. a review of an academic paper). This information on assessed oral presentation tasks were logged on an excel spreadsheet. Marking criteria and task briefs from subject programmes and EAP programmes were also obtained in this phase of the research. The task briefs contained information on the topic and format of the presentation, and in some instances gave top tips on how to approach the assessment. These internal documents were used to further understand how enacted performances may be shaped. A limitation of this stage of the study is that it has not resulted in sufficient understanding of how assessment formats and documents operate in

practice. However, the information garnered indicated the potential patterns and diversity in AOP practices within and across one institution: S1 analysis is discussed in Chapter 4. The data collected from the website survey of AOP tasks and marking criteria and task briefs was primarily aimed at aiding the design of the questionnaires for subject lecturers and EAP practitioners on the topic of AOP assessment.

3.6.2.2   Qualitative Questionnaires

In S1, in order to begin to develop understanding of AOP practice before entering the field sites, I wished to hear from HE practitioners with lived experience of AOP assessments. I decided that qualitative questionnaires would be the most effective and viable option with the given purpose, timescale and resources. I wished for the questionnaires to point to a range of avenues which may become the core focus of the project (see questionnaires in Appendices A and B). As well as this, I sought information on patterns and diversity of practice which could be used to compare with fieldwork data: I reasoned that through noticing similarities and differences across different contexts, I could better appreciate the characteristics of practice in the particular settings in S2 (see Appendix C for a detailed rationale behind the design of the questionnaire questions).

Conducting interviews was not deemed feasible in S1. This is mainly because interviews would have yielded vast amounts of data to analyse and require unreasonable time and resources of the participants and researcher, given the exploratory nature of, and emphasis placed on, S1 phase of this project. Murphy and Dingwell (2003) state that elicited texts such as questionnaires have the potential to generate data that resemble interview data. From my perspective, the use of predetermined questions read by respondents who have no contact with the questioner barely resemble the interview situation. However, I agree with the premise that the questionnaire is not limited to use as a tool for gathering quantitative data. I regarded a questionnaire as a useful initial method by which to gather qualitative data for exploratory purposes. Due to the exploratory nature of S1, I opted to design a questionnaire with predominantly open-ended questions. The open-ended questions provided the opportunity to receive richer data which give insights into the factors involved in phenomena that have received little attention in empirical research. The core reason for using such qualitative questionnaires was to gather preliminary insights into the practice of a variety of HE practitioners in a short time period. In addition, open-text boxes, rather than for example Richter scales, afforded the respondents the opportunity to provide individual responses which acknowledge the complex, context-dependent nature of practices.

Online questionnaires were used to reach participants at various locations in the UK quickly and to facilitate analysis as results were imported into NVivo (see Section 3.9 on data analysis and

Appendices I and J for examples of the S1 NVivo analysis). This style of questionnaire is not without limitations. First, the questions are pre-fixed; there are many areas in these questionnaires which would benefit from being explored further in follow-up questions. Second, the questionnaires with open-ended questions required a high level of engagement. In future, if using qualitative questionnaires, I would include fewer questions as this may result in fuller responses to some questions and higher completion rates.

I have described S1 methods, and now I turn to a discussion of S2 setting, informants and methods.

3.7   S2 Setting and Informants

In S2, I negotiated access to two field sites and spent time with one group on an EAP module on a year-long programme at each site. In this section, I share key details about the key informants, setting and assessment task. Further ethnographic details of the context and actors follow throughout the resultant chapters.

3.7.1   Informant Information

In FS1, I had 7 key informants -- 2 teachers and 5 students. There were 14 students in the group and Adam was the main teacher on the module. I observed lessons with Adam and the 14 students. Daniel was Adam's co-assessor for the AOP assessments for this group, and he had taught the students on different modules on the Pre-Master's programme.

In FS2, I had 5 key informants -- 2 teachers and 3 students. There were 10 students in the group and Georgina was the main teacher on the module. I attended lessons with Georgina and the 10 students. Tracey, the second teacher informant, was Georgina's co-assessor in the AOPs. She had contact with the students on other modules on the programme. Please refer to Table 6 below for the key informant details:

Table 6: FS2 Informant Details

Field site	Teacher/student	Pseudonym
FS1	Main teacher	Adam
FS1	Teacher	Daniel
FS1	Student	Henry
FS1	Student	Ivan
FS1	Student	Anna
FS1	Student	Fiona
FS1	Student	Ben
FS2	Main teacher	Georgina
FS2	Teacher	Tracey
FS2	Student	Marina
FS2	Student	Aimee
FS2	Student	Tara

3.7.2   FS1 Setting

FS1 is at a university in the UK. The group I spent time with were one of two classes studying on a module which was part of a Pre-Master's programme. In my very first meeting with Adam, the main teacher of the module, he recounted how the programme he teaches on is offered as a year-long or semester-two only course of study. Adam teaches a module, with a stated aim of improving English for master's level study. In the Pre-master's programme there are also content pathway modules in broad topic areas, which are general because the 11 students are destined for mixed disciplines, all of which cannot be represented with individual pathway modules. The syllabus is divided into 3 themes Teamwork, Analysis and Learning. Groupwork, reflective learning, critical thinking and argumentation and skills needed for data collection and analysis for research projects are part of the learning objectives. There were six assessment tasks in semester The AOP was weighted at 10% of the overall score in the semester and may form part of the final assesment portfolio of work; the students have another AOP task in the second semester whose mark will count towards the final assessment outcome should it be higher than the AOP in semester one. 20% of the overall grade for the Pre-Master's course was for speaking, with an assessed seminar as the second speaking assessment genre. The students complete the Pre-master's programme with the prime purpose of securing a place on a master's course

commencing in the next academic year. I spent time with the teachers and students during the first semester of study.

3.7.3   FS2 Setting

FS2 is also at a university in the UK. The teachers and students I spent time with were the only class of students enrolled on a module aimed at improving English for academic study on a year-long IFP. I was at the site during the latter part of their studies. The students were on the programme primarily in order to meet entry requirements in order to progress onto an undergraduate course.

Students enrolled on the course because they did not have A-level equivalent qualifications and/or the relevant language entry requirements. In the IFP there are also content modules which may not necessarily be closely related to the students' target disciplines. The groups consist of 8 students from mixed disciplines and therefore the content modules cannot be tailored tightly to the students' target disciplines. The content modules are flexible and give the students a 'taste of what is to come' (FS2, Georgina, IBP).

Module documentation states that the course helps students use English for academic purposes confidently in an academic environment. A coursebook is used as a guide and reading, listening, writing and speaking, vocabulary, grammar and academic study skills are taught using an integrated approach. The AOP assessment task is weighted at 25% of the IFP and contributes to the final assessment outcome.

3.7.4   FS1 and FS2 AOP Tasks

The AOP assessment tasks and rating discussion were the central events of interest and key focus of this study.

In FS1, the students had the opportunity to present twice in the year; the presentation which received the highest mark would form part of an assessed portfolio of work. In semester 1, the students were tasked with designing and distributing a questionnaire and reporting on the findings in an oral presentation in groups. For the assessed task, three students from another group on the module joined the group I was spending time with because the teachers wished to ensure the students in each group did not all share the same first language. For example, a student from South Korea would join a group of students from China. Adam wished to encourage students to communicate in English when conducting the research and preparing the

presentation. The students had delivered a practice presentation earlier in the semester but on a different topic to the assessed presentation and received feedback from Adam and their peers.

In FS2, the students chose a topic to research and present individually. Georgina, the main teacher on the module, checked the topic had enough academic mileage. Georgina encouraged the students to approach a topic in a way that could showcase their ability to critically discuss their content. The students had the opportunity to deliver a practice presentation ahead of the assessed one to peers and Georgina and to receive feedback on a script from Georgina. The students received intensive support in finding sources and extensive feedback from Georgina and peers at numerous points throughout the year.

In FS1 and FS2, the teachers had not received specific training in rating the AOPs on this particular module. In FS1, there were in-house standardisation meetings for pre-sessional and in-sessional courses over the preceding years. Adam was also an external examiner at another Pre-Masters Academic English programme. The marking criteria were based on pre-sessional marking criteria which had been developed within the EAP department and these went through Faculty QA procedures and had been shared with external examiners. Adam adapted the criteria and oversaw the marking standardisation for the module.

In FS2, there was no assessment training at the institution or EAP unit of which Georgina was aware. She recounted how she had learned to mark from discussing the process with colleagues. The criteria were developed by a team of EAP practitioners at the EAP unit. In FS2, there is an external examiner who covers the IFP and an EAP external examiner on the validation panel. Georgina adds that guidance or assistance from management is non-existent.

Further details of the tasks are in Section 5.2, including information about the rating scales.

3.8 S2 Procedures and Methods

This section gives an overview of the access negotiated to the two field sites, and the particular application of methods in this part of the research.

3.8.1 S2 Access and Integration

S2 involved gaining access to a class on modules which use AOPs as an assessment genre at different UK universities. I wished to be part of the social setting and explored ways to integrate into the sites so as not to be a wholly distant observer. I observed and documented practice on modules which were targeted, in part, at improving and assessing English and determining students' readiness to embark on an undergraduate or postgraduate taught course at a UK

university. I advertised the project via email with established contacts and via BALEAP -- an international forum for EAP practitioners. I approached gatekeepers at 4 institutions, which involved confirming participant recruitment processes with university Ethics committees. Due to the richness of data gathered, I chose to limit the field sites to two.

While conducting a study in a reasonably structured environment at the outset seemed to be a constraint on the ability to access a community, the very structured nature of the educational setting meant that it was easier to blend in and integrate to some extent into the setting very quickly (as mentioned in Section 3.5 above). The type of access granted and nature of possible interactions led me to 'develop creative methods that align to spatiality rather than remain restricted by sites' (Tsolidis 2008, p.281). My limited time in the community placed even greater importance on finding a research role (Walford, 2009) which enabled social interactions in the site of interest. I developed ways to become a more active member when appropriate rather than a 'peripheral member' (Adler and Adler, 1998) in the community of interest in a way that suits the setting and my disposition. As discussed in Section 3.5, being a peripheral member was appropriate and ethical in particular interactions so as not to compromise the validity and reliability of the assessment decisions being taken in the sites. Entering teaching and assessment contexts also entailed planning and meant fewer spontaneous interactions would take place. I took opportunities to become a participant in the classroom setting, to blend in by chatting with classmates, to have informal chats with teachers, and to keep in contact with key informants for discussions about interpretations of the data after the study. I have come to realise that becoming a credible partaker in people's social worlds in the research sites, like in all social situations, is a matter of gradual negotiation, pushing the boundaries of your comfort zone in some instances but remaining true to your own particular nature.

3.8.2   Staging of Methods

The choice of methods revolved around the central events of the AOP assessments. Therefore, I conceptualised the S2 fieldwork as having three stages. I sought to capture perspectives before and after the assessments to be able to trace any developments in thinking and actions over time, pre and post the event. How I view the staging of the methods is depicted below:

Before Assessment	Assessment	AfterAssessment
- Observations of lessons - Informal discussions with teachers and students - Audio-recorded interviews with students - Audio-recorded interview with main teacher assessor	Observations and video recording of presentation assessments	- Observation and audio recording of rating discussions - Audio-recorded interviews with students - Audio-recorded interview with main teacher assessor -Participant checks

Figure 5: Stages in S2 Methods

The application of the methods takes on a particular form on each course due to local organic practices. The same methods, at different intervals were conducted in FS2. The group is smaller in FS2, therefore 3 students were interviewed rather than 5.

3.8.3   S2 Methods and Rationale

This section provides details about the methods used during fieldwork and the rationale behind their application in this project. The methods discussed in Sections 3.8.3.1-3.8.3.6 are document analysis, observation, audio-recorded interviews, audio-recorded rating discussions, video-recorded presentations, and field notes.

3.8.3.1   Document Analysis

Documents such as the syllabus, marking criteria, handbooks, lesson materials, and assignment task briefs were sourced for analysis. This is often justified by claiming that student and teacher practices are reified within textual forms (Wenger, 1998; Tummons, 2010). However, while influenced by these documents, this study explores how teachers and students align and distance themselves from these texts. Prior (2008) advocates shifting our view of documents to focus on what documents do, rather than only concentrate on what they contain. The materials exert influence on thinking and actions (as discussed in Section 2.8.4 on posthumanist thought). For instance, the constructs described in the criteria were compared with what the teacher assessors and students expressed as important in lessons, interviews, informal conversations and rating discussions (see Chapters 6 and 7).

3.8.3.2   Observations

Observations of lessons, AOP assessment performances and rating discussions were conducted to be able to construct the social world of the EAP groups and events in their natural settings. The observation in situ provides thicker descriptions of practices. Being present at events entailed taking decisions as to my positionality and role in the interactions (as discussed in Section 3.5 and 3.8.1 above). In FS1 and FS2, I asked questions and gave feedback on practice presentations similar to what the teachers and students were doing. In FS1, I also delivered a presentation about being a postgraduate researcher after the other students gave practice presentations which further helped to become an active member of the group as I was seen as a fellow research student, a fellow presenter, and a fellow audience member in this class rather than purely an outside researcher and observer. I did however, not wish to 'go native' (Cohen, Manion, and Morrison, 2007, p.129) in these situations by taking on the role of a second teacher for instance and becoming heavily involved with helping students with work or by commenting on students' AOP assessment performances in rating discussions, so I could retain my main role as a researcher and focus on making field notes. However, there were episodes in which I took on roles that created closer proximity to the events or built rapport with informants. For instance, in FS1 a teacher asked if I could fill out an observation feedback form as part of a teaching qualification they were completing. I watched events as a peer teacher and as a researcher. In FS1, I gave a presentation about conducting research after the students had delivered presentations. This resulted in participating as a fellow presenter as well as observing events from a researcher standpoint. I believe such participation afforded a greater sense of shared experience. My identities as a student and EAP teacher were permitted to enter the fieldsites and I believe afforded greater affinity between myself and the informants and, consequently, the teachers and students seemed more willing to share their perspectives on events

I chose not to audio record the classroom sessions as this research does not focus on classroom instruction and interaction and I felt it would be unnecessarily obtrusive. Field notes were made and are discussed below in Section 3.8.3.6. I also observed the assessed presentation sessions and the rating discussions in the field. The rating sessions were audio recorded, and presentations video recorded but being present and observing in these settings proved crucial to getting an overall impression of the experience and richer interpretations of people to transport myself back to when conducting analysis and writing after the initial encounters.

3.8.3.3   Audio-Recorded Interviews

Two phases of audio-recorded intensive interviews (Charmaz, 2014) were conducted before and after the AOP assessment with the teacher of the groups and with the students who volunteered

in the early stages of fieldwork. These interviews allow the teacher and student interpretations of events to be accessed. Interview prompts were used and a copy provided to students before the interviews. See Appendices E-H for interview prompts. A reflection task was given to students to note down thoughts after the presentation assessment in preparation for the interviews after their assessed AOP.

Interviews were chosen as a method because they gave me the opportunity to ask students and teachers individually about their AOP thinking and practices and to facilitate understanding of the choices made in the AOP assessments and rating discussions not visible when observing practice or analysing talk. Charmaz (2014) believes that intensive interviews:

> give research participants a space, time – and human connection – to reflect on these events anew and to clarify meanings and actions while providing rich data that spark analytic insights (p.80)

Talking about teaching and assessment practices related to a specific group of students and institution can be a sensitive issue and so individual interviews were deemed appropriate, rather than a method such as focus groups.

In both the audio-recorded interviews and rating discussions, I used digital voice recorders. In the audio-recorded rating discussion, I made field notes, but the use of field notes in interviews were minimal as taking notes causes distraction (Walford, 2008). Two rounds of interviews were conducted with the key informant teachers and students. Intensive interviews were used which 'typically means a gently-guided, one-sided conversation that explores a person's substantial experience with the research topic' (Charmaz, 2014. p.56). In FS1 and FS2, first round intensive interviews began with open-ended questions and gradually more theoretically driven and focused questions were posed in order to explore nascent categories. I used lesson materials and marking criteria and slides to encourage the students and teachers to talk freely at length with minimal prompting at times – although some students required more clarification and prompting on occasions. The questions covered a range of topics and sought perspectives from interviewees without the inclusion of leading questions. The interviews before and after the assessment included some of the same questions so as to give participants chances to show any development, ambiguity and complexity in their thinking and practices. The questions in the pre-presentation interview pertained to the students' and teachers' biographic information and previous presentation experience, questions about the course instruction and assessment events and documentation, and their views about what makes a successful presentation on this course.

Prompts were provided (see Appendices E-H) to give students an opportunity to reflect on the issues before the interview and to put them at ease. While the topics were provided, the range of topics and unspecified direction from the researchers as to what to explore in FS1 and to some

extent in FS2 meant there was scope for the interview to be shaped by the students' and teachers' agendas. Similar topics were covered in the student interviews, but the questions asked and focus devoted to certain topics varied to a degree. I also used visual prompts to clarify questions and responses. For instance, when gauging views on what students would assess in AOPs I drew a circle on a piece of paper and asked students to divide it up and label it with marking criteria. This helped students to express views clearly and encouraged them to take time over the response.

Atkinson and Silverman (1997) warn against assuming that interviews link directly to the real experience as interviews are tainted by the participants' subjectivity. I would argue, in line with Husserl, that all behaviour is affected by subjectivity and the effect of the other's gaze. The researcher's account of reality is also subjective and one construction. The plausibility of the account has been strengthened by triangulating data from observation, interviews, informal conversations and documentation. Furthermore, attending to multiple perspectives on AOP performances and seeking to look in-depth into the language used and to the non-verbal communication and emotion injected into parts of informants' utterances has allowed me to ascertain when some accounts may be less theoretically plausible. For example, more nuanced understandings of students' views on the importance of linguistic features in AOPs were gained from triangulating data and tracing different statements of belief throughout the study. Resonance is a key quality of the interpretations which has also been important, and in building interpretations I have asked myself and key informants whether the constructions of practices ring true as well as seem plausible in member checks.

3.8.3.4   Video-Recorded Presentations

As a member of an audience and as a teacher assessor in the past, I have watched numerous presentations and with the passage of time presentation fatigue sets in and often all that remains after their delivery are the PowerPoint slides, some scratch notes on a piece of paper, and in most instances a hazy memory of the performance. Therefore, it was crucial to the study to obtain video recordings of the AOP events as data sets. Having copies of the video recordings was also crucial as FS1 teacher assessors played clips from these when assigning marks. Daniel in FS1 seems to appreciate the AOP recordings too – 'Gosh, these are useful, aren't they' (FS1, rating discussion), he remarked when he and Adam referred to segments when assigning marks. It is especially necessary for the purposes of this research because AOPs are such complex multimodal events where interaction between the slides, verbal and non-verbal aspects may be paramount to achieving successful communication. And if an in-depth ethnographic, interpretative account is to be achieved of how teachers and students make sense of such phenomena these aspects must be

captured and at least available for analysis. The decision-making processes have not been checked against the video-recorded AOPs, however, as I was not evaluating the teacher assessors' decision-making against the performances in this study.

The video recording of presentations is standard practice in some EAP assessments, therefore in FS1 and FS2 the use of a camera did not disrupt the natural setting. The participants consented to me having access to these recordings and the teachers agreed to send me the files via email. Studies which investigate AOP-related phenomena (Morita, 2000; Sundrarajun, 2007; Bell, 2010) have also included video-recordings of AOPs as one of their data methods. Analysis of the video data would have given an evaluative perspective on the AOP performances which might provide a different lens on the teachers' assessments and the constructs which underpinned them. The videos may be transcribed and analysed in future research.

3.8.3.5   Audio-Recorded Rating Discussions

Having audio-recorded rating discussions was particularly useful in capturing the teacher assessors' interpretations of AOP performances and for tracing their decision-making processes when they were engaged in reaching score decisions. The rating discussions offered a unique insight into the teachers' assessment thinking and behaviour. Gipps et al. (1995) point out that documentation is a means by which to access teachers' assessment knowledge. However, when decisions made are decided collaboratively through talk (as in FS1 and FS2), the recording of the rating discussion is a crucial means by which to access the black box of the qualitative and quantitative judging and marking that takes place in assessment decision making. The research is mostly an analysis of content related to decision-making rather than a conversational analysis. A conversational analysis could be conducted in future avenues of AOP research.

3.8.3.6   Field Notes

Field notes were used to document my observations of participants' actions, my interpretations of events and reflection on the research process. The notes 'facilitate a trail' (Kiely, 2001) of AOP assessment events and interpretations of them.

> Through recording an observation in your field notes and studying it, you may also gain fresh
> theoretical understanding and direction for your study (Charmaz, 2014, p.38)

Field notes were used alongside other methods to document information that is not included in audio and video recordings and as a means of capturing real-time observations often lost after the event. The field notes act both as a source of extra data and as a means to aid the understanding of data from other sources. In other words, the field notes help to transport myself back to the

settings I was in and to act as valuable sources of data both enriching the description and discussion and strengthening the plausibility of interpretation of events.

Notes were made during and after observations on the classroom layout, stages of the lessons, the interactions between the group members, my integration into the settings and the themes included in feedback given on AOPs. Feedback which could have been given on practice presentations was also noted down. During rating discussions, I jotted down parts of video clips which raters referred to during the discussion, and any notable non-verbal communication. I also made notes on ethnographic conversations outside the formal classroom settings. For instance, I wrote about how Georgina and Tracey, teacher assessors in fieldsite two, were discussing weaker and stronger students over lunch and how their previous encounters with students is admitted into formal assessment events.

Decisions made as to what to note down during interactions with participants were revisited throughout the study. The teacher in one site during a lesson commented 'oh you are writing a lot hehe'. I started to note key words and would revisit and add to these notes after the sessions so as to better blend in with the natural happenings of the setting. The reflections included in this section on the use of field notes have been made possible by using reflexive accounts of the research process written in such field notes and in memos related to these notes.

Field notes facilitated viewing events with an analytical eye. The next section expands on the analytic tools and processes enacted in this research.

3.9   S1 and S2 Data Analysis

In S1 and S2, I have used grounded theory as a guiding analytic framework. Figure 6 below represents the different stages in the collection and analysis of data in building a grounded theory. The processes, although depicted in a linear arrangement, are not applied in a fixed linear fashion.
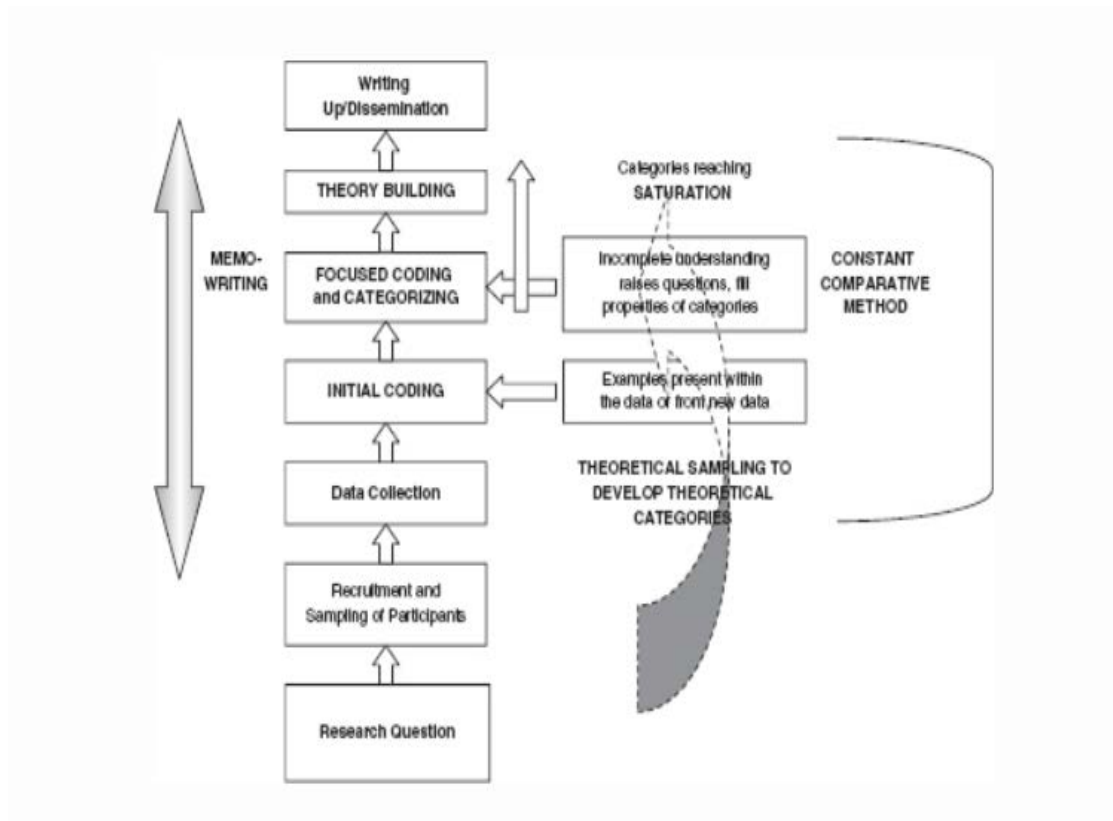
Figure 6: Visual Representation of a GT (Charmaz, 2014, p.18)

The analysis of S1 data used qualitative thematic coding and then CGT methods as a guide. When receiving documents and responses I first spent some time reading these to familiarise myself with them. I then made thematic codes on aspects assessed and then merged the aspects assessed into categories in a codebook. The design of the questionnaire was informed by the emerging codes and gaps in categories that arose from inductive and abductive reasoning. I imported the data from the questionnaires into NVivo to facilitate controlled and systematic data management and analysis. With the questionnaire data, I conducted line-by-line initial coding, focused coding and constant comparison methods and wrote memos. Memos were used to compare and analyse data from all data sets in S1. The memo-writing feature on NVivo was used to compare raw data from specific nodes and conduct incident-with-incident comparisons (See Appendices I and J for examples of the coding and memo writing analysis conducted on S1 data). These analysis steps engendered the identification of areas of interest to explore and bear in mind when conducting The Fieldwork Study.

Hadley (2017) states that a vast amount of studies do not conduct theoretical coding and therefore are using 'qualitative data analysis of theme which have been derived through using the methodology inspired by grounded theory' (p.119). In S1, I devised codes from the data which attempted to build a nascent typology of AOP assessment types, functions and considerations. These codes, although based on data, had become more abstract and theoretical in nature. This

was the first emergence of some theoretical codes which I would consider the relevance of in S2 fieldwork.

In study two, the analysis of data had two fairly distinct stages: precursory analysis stage during fieldwork and intensified analysis after fieldwork. Dornyei (2007, p.250), describes a period before coding when the researchers familiarise themselves with and engage with the data in its context 'meaningfully'. I believe that I went through such a process, to an extent, but I also decided to have a taste of coding while at the field sites as I found this also helped me to interact with the data and gain meaning that was not always visible from observation and field notes alone. I found parallels, and resonated, with reflexive accounts on the coding process in Hadley's (2012) PhD study. I too found the coding process daunting and it was difficult to find time to code during fieldwork, especially as a novice researcher new to grounded theory methods and analytic frameworks. I decided to do selected coding of extracts of data which I had identified as critical, due to the time restrictions imposed during fieldwork.

After the fieldwork had finished, I was fortunate to attend a 3-day workshop with Kathy Charmaz which involved hands-on experience of going through the analysis process. The training in 2018 came at a point when I had recently finished fieldwork and was transcribing interviews and rating discussions in full. The interviews and rating discussions were transcribed in intelligent verbatim style. I included overlaps, noticeably louder than surrounding speech, ancillary features such as long pauses, changes to speaker's tone and some body language (see Stylistic Choices for transcription conventions). After the workshop, I began to conduct line-by-line coding in a more systematic, rigorous and more confident manner thanks to this training opportunity. I realised that I had done some valuable analysis groundwork from the precursory coding process, but the codes were not sufficiently process-oriented and I had not drilled down to the heart of the actions in previous coding. The initial coding during fieldwork was of enough worth not to be disregarded. Some of the efforts in the pilot part of the coding contributed to the final interpretations of events in the field and it had been useful in filling emerging categories while in the field. Table 7 provides a clearer illustration of how the analysis processes were enacted during and after my fieldwork.

Table 7: Steps in S2 Analysis

S2 Analysis steps
Stage one: During fieldwork
1. Writing field notes throughout FS1 and FS2 data collection
2. Conducting line-by-line coding on documentation and field notes
3. Drawing out themes and focused codes in documentation and parts of field notes to inform lesson observations and interview prompts
4. Listening to interviews and writing notes on interesting episodes
5. Identifying key episodes in interviews and transcribing these
6. Line-by-line coding in interview and rating discussion transcripts of most important episodes in NVivo
7. Focused coding in NVivo
8. Identifying interesting leads and gaps in emerging categories from inductive and abductive reasoning
9. Application of constant comparison methods by analysing data in memos in NVivo
10. Theoretical sampling – questions in interviews and informal chats to gain information to fill categories and search for contrasting data and variations
11. Reference to theories in assessment literature
Stage two – After fieldwork
12. Full transcription of (central data sets) all interviews with teachers and students and the rating discussions
13. Line-by-line coding of transcripts and field notes from scratch (previous exposure to data during fieldwork and experimentation with codes was valuable)
14. Focused coding (See Appendix K)
15. Incident-with-incident comparison
16. Writing memos on data (See Appendix L)
17. Writing up in chapter form – analysing further and identifying gaps
18. Member checks with teachers to check interpretations and conduct theoretical sampling
19. A more focused survey of literature on relevant theories surfacing from the data analysis

The analysis process was iterative but the general progression in analysis involved initial coding, focused coding and memo-writing. The initial line-by-line coding fractured the data in order to drill down to the actions and processes involved. Each line was not necessarily coded, in some instances larger chunks of data clustered under one code. For example, a whole incident recorded in field notes, such as an ethnographic conversation, sometimes made up one code. From analysing the codes by subsuming smaller codes into larger focused codes, by discarding others and retaining initial codes which in fact began as larger categories, theoretical categories were constructed (See Appendix K). I found the coding process extremely fruitful in coming to know my data and to see the complexities in the processes at work connected to AOP assessments. The coding was an invaluable part of remaining true to my participants' words and not moving too far away from their worlds and interpretations in the construction of my narrative of the events. It was only in writing memos and draft chapters, however, that the analysis really started to take shape. I brought raw data into the memos with the list of open and focused codes in order to build larger categories (See Appendix L). In memos, codes then 'grow beyond mere means for sorting data and become processes to explore' (Charmaz and Mitchell, 2001, p.167). Building a memo on codes and raw data anchors ideas, and ultimately, balances evidence with theoretical argument. (Charmaz Mitchell, 2001).The emerging grasp of larger processes in writing memos then gave a greater ability to understand specific actions. Crucially, the time with students and teachers (the ethnographic dimension) facilitated the interpretation of data.

Writing memos fostered the identification of areas which may be underdeveloped and require more data and in-depth analysis. I conducted some theoretical sampling to provide fuller categories and richer analysis: for example, in including questions related to focused codes in the intensive interviews. In this way, grounded theory methods have involved inductive analysis but also abductive, as I thought of ways to be able to account for existing data. Two interviews were conducted and participant checks were conducted with the 4 teacher assessors. This enabled me to revisit topics and to include additional questions in second interviews and meetings after the main data collection had been completed. I also used worksheets developed by Gregory Hadley (see Appendix M) which use Böhm 's (2004) Coding Families adapted from Glaser, to further sharpen the visibility of the relationships between core processes and the conditions under which they operate and change, after having an individual GTM workshop with Gregory Hadley. Theoretical sampling was used in some instances in order to refine analysis, however this is not directed at achieving saturation but at 'theoretical sufficiency' (Dey, 1999). Dey (1999) asserts that grounded theories have categories suggested by data rather than categories saturated by data. Aiming for theoretical sufficiency encouraged me to identify inchoate categories which may require further data, while also acknowledging that it is near impossible to really know when you

reach the point of saturation, if such a point exists. As stated in Section 3.4, I recognise that I could have engaged in greater theoretical sampling and search for variation if I had collected data from additional field sites. I Plan to interrogate the processes and concepts identified in this project in future research.

The issues of methods and analysis have been covered here to provide a transparent account of the practice. This attention to my research practice continues in the next section, in which I share specific ways in which I have enhanced validity in the research process.

3.10   Validity and Ethical Considerations

As stated in the introduction to this chapter, the previous sections have served to justify the methodological choices taken in this study. This section offers a focused summary of the qualities and strategies used to enhance validity in this research. The positivist notion of validity has been progressively reworked or replaced (for example by Lincoln and Guba with 'authenticity' and 'trustworthiness' and Wolcott with 'understanding') by qualitative researchers. Determining validity, trustworthiness, authenticity or understanding ultimately involves evaluating the researcher's account with the phenomenon under study (Maxwell, 1992).The conceptualisation of validity is reworked relative to the philosophical position taken on what lies external to the researcher's account (Maxwell, 1992) and the purposes of the research (Brinberg and McGrath, 1985).

In grounded theory studies the concepts *fit, work, relevance* and *modifiability* are often used as criteria for evaluating the quality of a theory (Lomborg and Kirkevold, 2003). Fit refers to the how categories should be generated from and stay close to the data. *Fit* is interpreted differently depending on the ontological position of the researcher. I align myself with the constructivist grounded theorist Charmaz (2000): 'a grounded theorist constructs an image of a reality, not the reality – that is, objective, true, and external.' (p. 523) Charmaz emphasizes how 'consistent' the categories are with experiences under study. *Work* relates to the power the categories have in predicting, explaining and interpreting what is happening in the studied phenomenon. The *relevance* criterion refers to how well the theory reflects the core processes and areas of concern. *Modifiability* refers to how a grounded theory could be adapted when new data emerge. (Glaser and Strauss, 1967).

The blend of GT and ethnography in this research offers numerous strategies and techniques by which to heighten validity. Charmaz (2006) notes that 'the grounded theory method itself contains correctives that reduce the likelihood that researchers merely superimpose their preconceived notions on the data' (p.51). The filtering of theories with use of data and constant

comparison are built-in checks in grounded theory which enhance the fit, work, relevance and modifiability of the theory.

In addition to the core processes in grounded theory, I conducted member checks in order to gauge the fit, work and relevance of my interpretations. I talked through the emerging findings with the four teacher assessors and asked if the data and findings 'seem consistent' and 'ring true' with their experiences. In member checks on the emerging model of rating decision making, Georgina and Adam both indicated that they had not fully reflected on what they were doing or noticed particular processes but that my account seemed to make sense for them.

The ethnographically-oriented data gathering enhances the closeness to the studied phenomena. I observed, wrote field notes and audio recorded interviews and rating discussions and video recorded assessed AOPs. The resulting thick description enhances the fit or 'consistency' with the lived experience (Charmaz, 2000). Furthermore, S1 and S2 benefited from the triangulation of data from multiple methods. Drawing on multiple perspectives of events from multiple data sources and informants over time was used to construct a plausible interpretation of EAP AOP practices. The validity of the interpretations is greatly enhanced by including unfiltered voices of those engaged in AOP practices. I include verbatim raw data to evidence and make compelling and plausible theoretical arguments. The validity of an account cannot be determined by following particular procedures but by evaluating the researchers' judgements (Mishler, 1990).There are numerous excerpts of verbatim data accompanied with thick description of events so any interpretation is visible and justified. These steps endeavour to present as transparent, plausible and legitimate a rendering of events as possible.

The capture and processing of data also involves ethical decision making. The research process involved approaching the treatment of individuals and data with caution, care, thought – ultimately with humanity. The study involved being open about my research intentions, gaining informed consent to enter the worlds of teachers and students engaged in complex and sensitive assessment practices and ensuring the research process did not harm informants.

Ensuring the level of promised informant and institution confidentiality and anonymity was of utmost importance. Access was negotiated to institutions, departments, to groups and to individual lessons, assessment sessions and rating discussions. I gained relevant permissions from institution Ethics Committees and numerous key gatekeepers. I had initial meetings with teacher and student informants in S2 in order to present the character of the research and provide relevant Ethics forms. The informants in S2, as their identity is known to the researcher and other actors in the study, have been informed that only linked anonymity can be promised. All names from S1 documents and S2 data have been replaced using pseudonyms and I have endeavoured

to disguise other identifiable signifiers (Christians, 2005) to ensure as high a degree of anonymity as possible. Documentation is institutionally owned and therefore is unable to be published and therefore I can reflect the content only. Due to copyright issues, and to ensure promises to minimise the potential of participants being linked to their data, course documentation, notably the rating criteria, has not been published in full, but summaries have been included and key words have been quoted and paraphrased.

Developing ethical relationships with the informants in the study has also been a central concern. During fieldwork, I made every effort to ensure that my presence did not have an adverse effect on the student test takers' performances and the teacher assessors' marking practices. I attended lessons and practice presentations to minimise any harmful observer effect in the assessment events. I have also been careful not to influence unduly the programme practices which could call into question the validity and reliability of the assessments in particular. For instance, I provided no extra help to individual students on their assignments before assessments which would put any students at an advantage or disadvantage, and I withheld making comments on the summative AOP performances before the final score decisions had been reached. However, the reactivity of the informants to my presence was not viewed as a negative occurrence; at times it was maximised rather than minimised. For example, the informants shared that they had analysed the marking criteria and their AOP practices more closely than they had previously due to my research.

During the project, I endeavoured to critically reflect on my practices and behaviours as a researcher. At times, I would take part in events assisting the teacher and students. In others, I would maintain a helpful distance (as discussed in Sections 3.5, 3.8.1 and 3.8.3.2 above). The ethical dimension, as with the teacher assessors reaching score decisions in this research, is a fundamental component of treating people humanely. Resisting against the pressure to rush through the process and go through the motions to 'get the research done' has been vital in keeping the human concerns at the centre of decision making as a researcher.

3.11   Chapter Summary

The chapter has described, and accounted for, the research design and methodological decisions taken in this project. The case has been made for using a dual methodology: the compatible and complementary mixing of grounded theory and ethnography was outlined. Justifications behind the specific application of methods in S1 and S2 were provided. It is hoped that the transparent accounts of the research activity in this chapter has convincingly assured readers that this research has been conducted in a responsible and ethical manner.

The following chapter reports the data analysis of S1, The Survey Study, which was conducted to build a picture of AOP assessment practices before embarking on fieldwork.

Chapter 4   A survey of AOP Practices

4.1   Introduction

This chapter presents an analysis of S1 data. At the beginning of the PhD the aim was to investigate: 'What is happening in EAP AOPs?' This broad question, with a focus on actions and phenomena, is characteristic of a grounded, ethnographically-oriented approach. In Chapter 2, I clarified how, from the outset of the project, I conducted a scoping review of existing research on AOPs in Higher Education and acknowledged personal, professional and academic influences that have shaped this research. As well as these areas of influence, at the very beginning of this project, prior to S1 data collection, I had sub-questions for S1 subject module and EAP data (provided above in Table 3 in Section 3.3 and below at the beginning of Sections 4.3 and 4.4).

The S1 sub-research questions, devised at the early stages of this research, were broad in scope and were mainly focused on establishing the role of language and communication in AOP assessments. This project, as already touched upon in Chapter 2, broadened its focus to investigate the pedagogical and sociological processes in AOP assessment in order to encompass the complex integration of factors involved in shaping AOP practices, including language and communication. In this chapter, data which contributes to answering the questions posed at the initial stages of the project will be discussed. Such a discussion shows how I gradually became exposed to the complexities and perplexities involved in researching and conducting AOP assessment. Interwoven with the data analysis in this chapter are my reflections as a researcher on the analysis process. This stage of the project (S1) proved useful to being 'current about the experience or situation that you will be studying'(Charmaz, 2014, p. 59). The analysis of website survey data, documents and the questionnaire data have shed light on subject module and EAP module AOP formats, the criteria (within the rating scales) used in AOP tasks, rater expectations and perceptions on the role of language and communication in AOP assessments. I analysed data by coding documents manually and coding questionnaire data in NVivo. I imported manual codes into NVivo and wrote memos to further refine categories (See Appendices I and J for examples). Table 9 below gives details on the achieved data set in S1.

Table 8: S1 Achieved Data Set

Method	Key details
Website survey of subject module AOP tasks	Webpages, 1 institution
Subject module task briefs and criteria	6 different modules, 1 institution
EAP module task briefs and criteria	7 different modules, 4 institutions
Questionnaire for subject lecturers	45 full responses, 8 institutions (see Appendix D for disciplines represented)
Questionnaire for EAP lecturers	14 full responses, 11 institutions

The raw data, analysis and reporting of data from S1 is almost exclusively qualitative (as discussed in Section 3.6), apart from some reporting of the quantity of responses which shared a particular theme and numbers used to represent scores in rating scales. Documenting the amount of questionnaire responses on a particular theme points to a theme's prevalence within the data set. However, it is not intended to indicate the salience of themes, as I reason that one response could prove to be of significance in local contexts in which individual HE practitioners or documents operate, as well as apply to other contexts even if not detected in S1 data. The survey data has some basis for generalisation however in that it captures typicality and the salient data provides insights into telling phenomena which could be identified in future research in other contexts.

I begin the chapter with information on AOP task formats on subject modules and EAP modules at UK universities from analysis of all S1 data sets (in Section 4.2). This sets up the analysis of subject module data in Section 4.3 and then EAP module data, in Section 4.4. Within the subject module data section (Section 4.3), I lay out key codes which emerged in S1. I cover key AOP functions (Section 4.3.1), the emphasis on content and delivery (Section 4.3.2) and perceptions on language and communication in AOP tasks (Section 4.3.3), before summarising the analysis of subject module data (Section 4.3.4). The EAP module data section (Section 4.4) contains the core analysis which has been divided into categories: EAP AOP task details and functions (Section 4.4.1), the areas assessed in EAP AOP marking criteria (Section 4.4.2), the role of language and communication in EAP AOPs (Section 4.4.3), treatment of language use (Section 4.4.4), and difficulty in noticing language use in AOPs (Section 4.4.5). A summary of the EAP module data analysis is then provided (Section 4.4.6). The chapter closes with a summary of the key findings of S1 data analysis and how they paved the way for S2 fieldwork (Section 4.5).

The next section demonstrates how data on subject module and EAP module AOP formats points to issues which are important to explore in The Fieldwork Study into AOP practices.

4.2   AOP Task Characteristics in UK HE

There is a lack of research which clearly documents current diverse and similar task types referred to as academic oral presentations (as discussed in Section 2.6). The analysis of the website survey of AOP tasks, marking criteria and SLQ data demonstrates that a range of AOP tasks take place at UK universities. Information about task format and requirements indicates that there are an array of aspects which need to be considered when developing a closer understanding of AOPs. The following table has been created which maps possible format features, which shape AOP practices. Table 10 is not an exhaustive list as it is entirely empirical, based on data from both subject module and EAP module data on AOPs in the achieved data set, which is by no means representative of all AOP assessment tasks in UK HE. Difference and similarities between subject module and EAP module AOP data will be highlighted later in this chapter.

Table 9: AOP Task Format Characteristics

1.	Assessment by	teacher of students; mixed-panel; peers; self-assessed; teacher who does not know students
2.	Presenter	individual; paired; group
3.	Assessment type	formative; summative; blend of formative and summative
4.	Stages assessed	preparation; delivery; post-delivery
5.	Audience knowledge	expert; a degree of knowledge assumed; non-expert
6.	Who is the audience	live audience; watching a recording; panel of academics and/or industry experts; teachers; peers; imagined audience e.g. a business client
7.	Raters assess as	a teacher; an examiner; an imagined audience; an industry expert; a peer
8.	Aspects assessed	content skills; research skills; communication delivery skills; interpersonal skills; group work skills; ability to assess peers; ability to self-assess; linguistic skills
9.	Q&A session	a crucial part of assessment; assessed but not exploited; present but not assessed; not included

The data from S1 provided an insightful snapshot of the diversity inherent in AOP assessment tasks in HE. For ease of reference and presentation, the format characteristics in Table 10 have been provided in categories. AOP tasks and practice may fall into numerous categories in each of the sections in Table 10. For instance, 'stages assessed' may include an AOP performance and a reflection on the performance in the form of an assessed written report (hitting both 'preparation' and 'delivery' categories in the 'stages assessed' section in Table 10). The fieldwork analysis in S2 shows greater complexity in AOP assessments: practices often move between and/or apply simultaneously to numerous categories in the sections outlined in Table 10. This signifies that the survey data touches on the complexity of AOP tasks, but that the integration and complexity is further illuminated and intensified in S2 analysis (see Chapters 5, 6, and 7). I share further examples later in this section.

Conceptualisations as to what is identified as an AOP were shown to be fairly fluid in S1 data. A prepared monologue is soundly established as a core property of an AOP performance; a dialogic, unplanned question and answer phase was a feature which was less ubiquitous, according to the S1 data from HE practitioners. In S2 data also, the extent to which a question and answer session was exploited varied between individual presentations. These findings illustrate how the label AOP encompasses wide-ranging practice and the difficulty in pinning down privileged properties and committing to a description and shared understanding of the construct of an AOP (as discussed in Section 2.6, and see Section 8.2 for a discussion on the construct of the AOP genre in light of this project's findings).

The website survey, document analysis and survey responses in S1 pointed to assessments which required diverse approaches to rating. The rater may be the teacher of the student test takers or take on an imagined role, such as that of a business client. In the survey responses, the picture of the possible rater roles is somewhat nuanced but, again, is less rich than the in-depth, ethnographic exploration of raters' AOP practice. In S2, Adam, the teacher in FS1, describes how he moves between different spaces. He marks as the teacher who has spent time with the student test takers and an examiner judging the students as an outsider who may not have the luxury of prolonged and close interaction (see Sections 5.3-5.7 and Chapter 6 for a fuller discussion of rating roles and processes).

The categories identified in S1 analysis acted as comforting anchors with which I entered the field sites, better equipped to navigate the diverse mix of rating roles, strategies and behaviours I would encounter. Some may argue these categories could be too restrictive and that entering the field site without such crutches encourages freer thinking. However, in this case, the questionnaire aimed to produce grounded codes used to indicate the myriad of possible

manifestations of AOP tasks in order to furnish my imagination for the complexities and perplexities to come. Furthermore, the S1 data provided points of comparison with which to better understand the particular make-up of practices in individual field sties.

The S1 data demonstrated that AOP rubrics on subject modules and EAP modules contain a range of criteria. When analysing marking criteria and responses to SLQ questions pertaining to what is assessed, the following focused codes were generated: content skills, research skills, communication delivery skills, interpersonal skills, ability to assess peers, ability to self-assess. 'Content skills' refers to a test taker displaying ability to describe and/or evaluate information which was covered during instruction or from new sources. 'Research skills' encompasses abilities tied to locating information and conducting research projects which involve application of research methodology. 'Communication delivery skills' identifies aspects crucial to the AOP delivery such as language use (oral and written on slides) and non-verbal communication. 'Interpersonal skills' represents such areas as not dominating a question and answer session and creating a positive rapport or connection with the audience, which admittedly bleeds into the communication and group work skills categories. 'Group work skills' cover aspects, beyond the general interpersonal skills, which are required when working specifically in a team rather than communicating to a group of people, such as specific organisational tasks or aesthetic considerations when compiling slides jointly and shared monologues. Applying fixed codes to the marking criteria proved problematic and there were instances of clear and grey areas of overlap. There are difficulties distinguishing content and research skills, for instance, as researching a topic for a presentation could be considered part of demonstrating content knowledge but also a research skill. Creating codes for aspects assessed from S1 data prepared me for the challenges inherent in attempting to crystallise what is assessed into neat categories.

The difficulty lies in establishing and presenting, in a coherent and helpful manner, what factors into AOP decision making. While a close look at practices show categories to be fuzzy and entangled, the act of deconstructing AOP phenomena into such discrete building blocks serves to illustrate the variety of influences on practice and, from my reading, rather than reifying the practice, inferred the idiosyncratic realisations of AOPs in HE. The one constant which leapt out from mapping the terrain of AOP assessment in S1 was that AOP assessments in UK HE were presented as a versatile genre called into the service of HE practitioners with varied agendas.

The aspects assessed in the AOP tasks linked in part to the purposes of the assessment. These purposes were identified as related, or unrelated, to the module objectives by HE practitioners in some responses (see Section 4.3.1 below). While the categories in the table have been identified in both subject module and EAP module AOPs, the next sections look at subject module data and

EAP data separately due to a reported difference in assessment purposes. EAP modules, which feature in S1, included language in the criteria and assessed test takers' readiness to function on subject modules conducted through the medium of English; while the subject modules were primarily assessing disciplinary knowledge and skills and wider graduate attributes. There is overlap, however, between the functions and AOP characteristics on subject modules and EAP modules, which will be highlighted in the following sections.

4.3   S1 Subject Module AOP Data Analysis

EAP courses are tasked with preparing students enrolled on them for the academic realm and therefore the practices on subject modules have been explored in S1. The data and focused codes which have surfaced from subject module data informed the future stages (the EPQ and S2 fieldwork) in this project. The collection and analysis of S1 module data was loosely guided by the set of following sub-research questions:

S1 subject module data sub-RQs

1) What are the purposes of AOP assessments on subject modules?
2) What criteria are used in subject module AOPs?
3) What language use is viewed as appropriate in subject module AOP assessments?
4) What challenges do students and teachers face in AOP assessments on subject modules?

In the next sections (4.3.1-4.3.4), I present an analysis of the S1 module data which addresses these questions.

4.3.1   Tasks and Functions

An analysis of SLQ data showed that the AOP assessment on their modules is regarded by subject lecturers as fulfilling a number of functions, addressing S1 subject module data sub-RQ1: 'What are the purposes of AOP assessments on subject modules?' The following focused codes were identified from SLQ data pertaining to the purposes of the AOP assessments: determining whether specific module learning objectives have been met; developing academic skills not possible in other (namely written) genres; affording students the opportunity to develop skills commonplace in their future careers; acting as an engaging and more enjoyable means of building rapport in the group; facilitating the marking process as quicker to mark than essays; fostering engagement with the module content; and diversifying the assessment methods to strengthen the fairness and validity of a module's and/or programme's assessment procedures.

The AOP was presented as a versatile assessment and learning tool due to its interactive dimension and the transferable skills it involves and develops. The need to master various skills when delivering an AOP was viewed as both complementary and detrimental to students'

learning on specific modules. 3 of the 45 respondents stated that the AOP was not a particularly useful method in terms of assessing how students had met the module learning objectives, while other lecturers regarded the AOP assessment as aligning with the learning objectives at the same time as developing key transferable skills. SL 33 (Music) explicitly mentioned that tensions emerge when trying to develop or assess graduate attributes by posing the question of whether subject lecturers are experts in these areas. SL 33 has grappled with the focus put on delivery and whether to assess it. The shift to active learning and development of successful oral communicators in HE has ramifications for the oracy demands placed on staff and students in HE (see discussion in Section 8.3). The emphasis which lecturers place on content and delivery is an important issue to consider and we turn to this next.

4.3.2 Content and Delivery

The criteria, task briefs, website surveys and questionnaire responses revealed HE practitioners often work with a content versus delivery distinction. The content referring to *what* is contained within the presentation and the delivery is *how* this content is communicated. The emphasis placed on content and delivery was found to vary in AOP assessments in the marking criteria and SLQ data. The following responses to question 3.2 (see Appendix A) demonstrate varying views on the importance of content and delivery. Content may be the priority (SL 20), delivery may take precedence (SL 26) or equal emphasis may be placed on content and delivery (SL 36):

> Content is first.
> Speaking clearly, slowly, inclusively is second.
> Both are essential for an A grade.
> SLQ, SL 20, Sociology
>
> Delivery & Timing, followed by content and visual aids.
> The second two are easier to improve, but the first two are in some ways totally critical
> to something being well received at a conference, interview, briefing, etc
> SLQ, SL 26, Archaeology
>
> 1) The quality and depth of information presented.
> 1) The speed and clarity of oral and/or visual material.
> 2) Audience engagement.
> SLQ, SL 36, Urban Studies

The data indicates the range of skills AOP assessments may involve, which addresses S1 subject module data sub-RQ2: 'What criteria are used in subject module AOPs?'. It has proven problematic ascertaining whether aspects mentioned by subject lecturers are assessed or developed in AOP assessments, based on the questionnaire data. And, in fact, this proves to be problematic in S2 when rating decision talk itself is observed, captured and analysed (discussion to follow in Chapter 6 and Section 8.6). The aspects -- assessed or developed -- have been grouped together into the following categories: content skills, research skills, communication

delivery skills, interpersonal skills, ability to assess peers, and ability to self-assess. As pointed out in Section 4.3.1 above, the aspects assessed may tightly cohere with the module learning objectives or be guided instead by other factors such as promoting particular graduate attributes. The AOPs may be used as a means to assess non-delivery aspects, oracy skills being purely the product of the task, or oracy may be the target of instruction.

One out of the 45 respondents explicitly stated that they assess content only in the presentations and presentation skills are not assessed. The vast majority of SLQ responses, however, indicated that SLs are concerned, albeit to a minimal degree at times, with how content is communicated. Oracy skills may become an explicit or hidden part of the assessment (Doherty et al., 2011; see Section 8.3 for discussion). Therefore, the SLs' expectations regarding the use of language as a communication tool is interesting to explore in the next section and contributes to answering S1 subject module data sub-RQ3: 'What language use is viewed as appropriate in subject module and EAP module AOP assessments?'

4.3.3 Language Use and Communication

In responses to the SLQ question posed as to what constitute the most important features of an AOP assessment on one of the SL's taught modules, content-related aspects were the most mentioned feature. However, three lecturers mention features of verbal communication as salient characteristics in communicating the content such as 'the ability to speak naturally' (SL 15, Sport medicine), 'clear communication of ideas' (SL 18, Education). One lecturer mentioned grammatical correctness as an important feature (SL 25, Animal Welfare). Although three is a relatively low number, the question was open-ended and when asked specifically about instances when language affects the AOP delivery and whether feedback is given on such instances a greater number of responses cited aspects of verbal communication deemed crucial to the AOP delivery, according to SLQ respondents.

I was dubious as to whether there would be much engagement with question 3.5 pertaining to language use in AOP assessments (see Appendix A), however the lecturer responses reflect the relevance of oral and written communication in AOP assessments and justify a sustained exploration into its impact. There were five lecturers out of the 45 who stated that they have never given such feedback on language to date. Four lecturers stated that they rarely give feedback; two of whom stated they tend to give feedback only on poor use of slides. On the other side of the continuum, four lecturers stated that they had given feedback on all of the aspects mentioned in the prompt (use of text on slides, register, pace, grammar, word choice, pronunciation) in their experience assessing AOPs.

> Multiple students have had all of the above. Yes, feedback is provided on these aspects are they are important for delivering an effective talk
>
> SLQ, SL 11, Genetics

The remaining lecturers had other experiences with language use in AOPs. Pace -- speed of speech -- was the feature which the highest number of mentions (fifteen). Eight lecturers stated they had given feedback on pronunciation and register, while eight mentioned vocabulary-related aspects. Some respondents did not provide specific examples and therefore did not give a clear indication of what language use is valued in the AOP. Other responses to question 3.5 pointed to potential motivations behind providing feedback on language use: inability to follow content or the impact on the ease and style of delivery of the content.

The perceptions of linguistic challenges faced by first and second language users of English were explored in the questionnaire, which partly addresses S1 subject module data sub-RQ4: 'What challenges do students and teachers face in AOP assessments on subject modules?' In answer to SLQ question 3.5 (see Appendix A), some SLs stated that feedback tended to be similar for students with various linguistic repertoires. Responses to questions about allowances (question 3.4 in Appendix A) made for first and second language users showed that some lecturers believed second language users may face additional challenges such as longer preparation time and heightened levels of anxiety. However, the responses also showed that lecturers believe that the similar challenges are shared by all students irrespective of being a first or second language user. Difficulties which arise depend on numerous factors (e.g. educational background, language background, cultural background and disposition). The challenges overwhelming identified by lecturers in the SLQ for all students were nerves and structure. In the questionnaire, I used terms native English speaker and non-native English speaker (see 9.2 on limitations for details on the quandaries faced in using the terms native and non-native English speaker). In regard to lecturer challenges, many SLs reported none, and the issue of rater consistency and bias featured most prominently. The perceptions and possible treatment of linguistic features of AOP performances and non-linguistic features of communication in subject module AOPs would be a fascinating area to explore in ethnographic research in the future (see further research avenues in Section 8.3).

4.3.4   Subject Module Analysis Summary

Section 4.3 has summarised the data from the website survey, marking criteria and task briefs from subject modules and SL questionnaires. In so doing, it has shown such survey methods to have been instrumental in informing this project by unearthing potential diverse and shared characteristics of, and issues involved in, AOP practice. The analysis demonstrates the wide use of tasks conceptualised as AOPs across institutions and disciplines; current varied AOP task formats;

the range of aspects which may be assessed or developed; SL perceptions on language use and communication in AOP assessment; and challenges faced by students and teachers.

The data shows that AOPs are used in HE as a multipurpose assessment and learning tool due to the affordances of the genre. Practitioners often use AOPs to develop discipline specific skills covered in the module and in some cases additional graduate attributes such as communication skills, if these are not already part of the module objectives. In some cases, the graduate attributes are targeted in the AOP assessment to the detriment of alignment with the module objectives, according to lecturers.

The analysis indicates that delivery -- the verbal and non-verbal communication during the presentation -- is viewed as important in AOPs for a number of lecturers. Verbal communication and written language on slides are viewed as crucial elements to consider when evaluating the clarity of communication of content in some cases. The amount and nature of feedback given on language in AOPs varies, however pace of speech featured most prominently and lecturers listed linguistic and non-linguistic challenges which were predominantly applicable to first and second language users according to subject lecturers. This empirical data evidences the importance of considering the role of oracy in AOP assessments in UK universities (see discussion of findings and implications drawn from data collected on subject module AOP assessments in Sections 8.2 and 8.3).

The subject module documents and questionnaire data provide insights into stated practices in UK HE for the purposes of informing an investigation of EAP AOP practices. From undertaking the subject module data analysis, the resultant firing of synapses created a wider nuanced landscape of AOP assessment. I established focused codes which aided me in revising loose research questions to then take on the task of grappling with the question: What is happening in *EAP AOPs?*

The subject module data informed the design of the EPQ and revised the sub-questions which loosely guided the EAP strand of S1:

S1 EAP sub-RQs:
1) What are the purposes of EAP AOP assessment?
2) What criteria are included in rating scales designed for use in EAP AOP assessments?
3) What emphasis is placed on linguistic and non-linguistic factors in EAP AOP assessments?
4) How is language use treated in EAP AOP assessments?

The documentation of AOP practices on subject modules that has been achieved gives a useful reference point for investigating practices on EAP modules and programmes. Such a reference point is especially valuable as EAP often seeks to reflect practices in the target language use

domain in some form. The next section documents and discusses EAP AOP practices, and particularly shines a light on those areas of interest which served to inform S2.

4.4   S1 EAP Module AOP Data Analysis

In the second phase of S1, I obtained EAP AOP marking criteria and task format information. This was achieved by sending out requests for documents and the online EPQ to EAP practitioners via the BALEAP mailing list. In the next sections, I share analysis of data regarding EAP AOP tasks and functions, the criteria in the rating scales, the emphasis put on language and the treatment of language use. This section (4.4) finishes with a particularly salient theme – the difficulty in noticing language use in AOPs.

4.4.1   Tasks and Functions

Task briefs and criteria obtained through communication with EAP practitioners via BALEAP and the information shared about the AOP tasks in answer to question 2.1 in EAPQ (see Appendix B) demonstrate the range of EAP AOP tasks and functions at UK universities and answer S1 EAP module data sub-RQ1: 'What are the purposes of EAP AOP assessments?' Table 11 below gives EAP AOP task details from four different institutions which illustrate diversity of practice. It includes whether the AOP tasks take place on an in-sessional or pre-sessional course, whether the students present individually or in groups, the timing allocated and the task description. All of the AOPs are described as summative, therefore contributing to the final assessment of the students on their respective courses.

Table 10: A Selection of Diverse EAP AOP Tasks

in-sessional - individual - summative - an oral summary of the main points of their essay and answer questions on it. *Source: EAPQ*	in sessional summative group presentation - 5 weeks to research a topic and create an academic presentation using sources. 5-7 speaking for each student plus time for questions *Source: EAPQ*
Summative group presentation on pre-sessional course, task to reflect on how course reading has shaped their thinking and essay writing *Source: EAPQ*	Pre-sessional, Individual, Summative Present a critical summary of a book article for an essay. 6-8mins +2-3mins questions *Source: Criteria sourced via email*

The findings from task information gathered contributed to the AOP task considerations which have been used to map AOP task characteristics (see Table 10 in Section 4.2 above). The AOP is operationalised in various ways so that it has the ability to fulfil multiple purposes and functions. In response to question 3.1 in the EAPQ (see Appendix B), EAP practitioners outlined numerous reasons why AOPs serve as a useful assessment task on the courses on which they have taught. The AOP tasks develop confidence in speaking in front of peers and teachers; develop and assess non-linguistic transferable skills related to content and presentation skills; prepare students for similar AOP events on, or beyond, their academic course; diversify assessment methods on the course; develop and assess interpersonal skills such as teamwork; provide students with a recognizable format; facilitate the marking process on the course as they are easier to assess than essays; encourage students to present information in a succinct and engaging way; facilitate the assessment of monologic and dialogic speaking (depending on how the Q&A part is exploited); give students the opportunity to practise listening skills; and provide an enjoyable and useful formative learning opportunity.

While the comments in the EAPQ were on the whole positive, an EAP programme leader contacted me via email and cited reasons such as presentation fatigue, lack of interrater reliability and presence of group think in marking as ones which contributed to dropping this task type from the EAP programme assessment methods at their institution. The only distinct difference overall between the diverse ways in which AOP task types are regarded as useful by EAP practitioners compared to subject lecturers are, unsurprisingly, the predominance of the language-specific related functions. In other aspects, including developing communication skills and assessing content skills, EAP AOPs seem to closely mirror subject module AOPs. A notable departure is that group work and interpersonal skills feature less prominently in EAP AOP data. The next section examines data which details what assessment criteria feature in rating scales used in EAP AOP tasks, addressing S1 EAP module data sub-RQ2: 'What criteria are included in rating scales designed for use in EAP AOP assessments?'

4.4.2   Rating Criteria

EAP practitioners commented on the integrated nature of the AOP assessment despite external UKVI requirements, for some institutions, which state the need to assess the four skills (reading, writing, listening, speaking) separately. While developing language skills is revealed to be a key objective in some AOP assessments, honing other skills necessary for the target academic environment is presented as fundamental to the rationalisation behind the use of the AOP as an assessment tool for most EAP practitioners in the questionnaire.

The EAPQ data analysis shows that criteria in the EAP AOP rating scales include content skills, research skills, linguistic skills, communication delivery skills, and group work skills. The naming of categories and what they involve in itself varied among responses. The conceptualisations of the constructs would need to be more fully understood by accessing marking practices in situ. Furthermore, it is important to ascertain whether aspects related to these purposes (e.g. group work skills, listening skills) are stated as aspects that are assessed or purely developed for formative purposes in assessments by individual practitioners. These areas were explored in the ethnographically-oriented study (S2) and the problematic nature of distinguishing what contributes to score decisions is elaborated on in Chapter 6.

Once I had coded the criteria, task briefs and questionnaire responses for aspects which were stated to be assessed, I then compared these to determine what was included, and not included, in individual assessments. I started then to conceptualise the purposes of the AOPs and began to endeavour to map out different types of AOPs which may exist on an EAP course. This typology was developed using primarily inductive reasoning and aimed at raising awareness of the possible uses of AOPs. Table 12 contains the working conceptualisations of three broad types of EAP AOPs which acted as a loose framework for conceptualising what was, and what was not happening, in AOP assessment tasks.

Table 11: Typology of AOPs

Linguistic skills assessment
Linguistic and academic skills assessment
AOP communicative performance assessment

A small number of responses (discussed in 4.4.3 below) in the questionnaire demonstrated that only linguistic skills featured in marking criteria, therefore generating the 'linguistic skills' assessment category. Other rating scales included linguistic features such as 'grammar' and also 'research skills' and 'content skills' such as 'use of sources'. These responses generated the 'linguistic skills and academic skills assessment' code. Another set of rating scales covered a vast array of criteria needed to deliver an AOP such as linguistic features, academic skills and presentation content and preparation skills, thus making 'AOP communicative performance assessment' a necessary category for reference.

In S2, I gathered richer data through interviews and rating discussions, an analysis of which evidences that criteria may not reveal what is valued or contributes to score-reaching decision making (see Chapter 6 and discussion in 8.4). The factors influencing decision making are

numerous and branch away from the skills outlined in the typologies set out from S1 data (see Chapters 5 and 6). However, these categories were a useful starting point and served to illustrate the potentially stark contrast between what is contained in marking criteria, practitioners' perceptions of what is assessed and what in fact is shaping rating practices in S2.

From describing the different types of AOP tasks, it is worth exploring in depth the emphasis placed on, and relationship between, linguistic and non-linguistic features in EAP AOPs. The preceding sections address S1 EAP module data sub-RQ3: 'What emphasis is placed on linguistic and non-linguistic factors in EAP AOP assessments?'

4.4.3 Role of Language

Data from the EAPQ points to the existence of strong and weak senses of performance assessments (McNamara, 1996: see Section 2.8.2). EAP practitioners may emphasise the linguistic elements in the criteria (weak sense of performance) or non-linguistic aspects (strong sense of performance).

EP 2 explicitly stated that 'language' was not included as a distinct category in the marking criteria in the AOP they had assessed but may fall under the category of 'delivery'. Criteria used by EAP practitioners 9 and 10 prioritise language use (see data extracts below), fitting in the 'linguistic skills assessment' category (in Table 12 above), with the criteria used by EP 9 purely focusing on linguistic elements. In contrast, criteria used by EPs 3, 4, 5 and 6 allocate around 50% or less of the weighting to linguistic features, falling most prominently into the 'AOP communicative performance assessment' typology (in Table 12 above):

Equally weighted: grammar, vocabulary, interaction, pronunciation, fluency

> EAQ, EP 10

40% linguistic features   20% Spontaneity in Responses   20%Presentation Skills   20% Presentation Content and Preparation

> EAPQ, EP 5

The EAP AOPs on the whole differed from subject module AOPs, arguably unsurprisingly, on the emphasis they put on linguistic features. However, when directly comparing particular AOPs, it was found that an EAP AOP in one context could place less emphasis on oracy skills than an AOP assessment on a subject course. Therefore, it cannot be assumed that subject module AOPs are

primarily assessing content-related aspects and that EAP AOPs primarily assess the student's ability to meet an entry language requirement.

Some EAP practitioners view EAP modules as language modules and therefore would reflect this in the weighting:

> We have had many discussions in our team about the weighting and keep coming back to the argument that the module is a language module and the weighting needs to reflect this.
>
> EAPQ, EP 7

In contrast, there are EAP practitioners who feel that it is not desirable to have instances when a student passes an AOP assessment due to strong language proficiency when their development of academic skills has been weak. EP 5's comment suggests that the weighting should centre on task fulfilment rather than language use and EP 3 believes content takes precedence:

> It is important to have criteria and an overall mark which reflects the task. If a student can speak well but has produced a poor presentation and the criteria is weighted 50% lang 50% presentation skills it can save their mark and so outsiders may think they are good at oral presentations when the reality is they can just speak well.
>
> EAPQ, EP 5
>
> Students with little academic content but good pronunciation and grammar will still manage to scrape a pass.
>
> EAPQ, EP 3

These comments from EP 5 and EP 3 compared with EP 7's comments highlight the different purposes assigned to the use of an AOP. The AOP can be used primarily to assess the test taker's ability to deliver the communicative act - the AOP, or the AOP task can be used namely to assess particular skills through its use. Deep-seated questions and tensions emerge here as to the role of the EAP practitioner. I remember as a novice EAP practitioner being struck by how little of my own EAP instruction focused on language instruction. The construct of EAP as discussed in section 2.4 is constantly evolving and EAP practitioners are continuously faced with decisions as to how to develop language and overall readiness for academic study (discussion of the EAP construct and implications for EAP practitioner identities and roles is covered in Section 8.4). The choice of aspects assessed and weighting allotted is an area to be explored by considering the specific make-up of particular courses, their learning and assessment objectives and individual rater and presenter priorities. These are aspects explored in S2 fieldwork in analysis Chapters 5, 6, and 7 and Sections 8.2-8.5.

The EAP practitioners were asked what they deemed to be the most important features of an AOP on the EAP module they have chosen to write about in the EAPQ. EP 4 (see data extract below) explicitly states that language use only comes to the forefront of decision making if it affects the delivery or content elements in the task at hand. EP 12 also shares this view but highlights that basic repeated errors are problematic.

> Content and Delivery - these relate to task achievement; Use of English tends to be mentioned only when it becomes relevant (either positively or negatively) to supporting Content and Delivery.
>
> EAPQ, EP 4
>
> Fluency is definitely important, as is varied vocabulary and style.
>
> Grammatical errors are not too much of an issue as long as they are not basic, repeated errors constantly.
>
> EAPQ, EP 12

These responses show that there tends to be a communicative focus in the treatment of language. EP 4 views the AOP as an assessment of a communicative act and this finding contributes to the notions of what constitutes the construct in EAP AOP assessments (further explored in Chapters 5 and 6 and Section 8.4). The next section explores data analysis related to S1 EAP module data sub-RQ4: 'How is language use treated in EAP AOP assessments?'

4.4.4   Treatment of Language Use

Whether the purpose of the AOP lies in assessing language proficiency more generally or the test taker's ability to deliver an AOP as a communicative act, or combination of these approaches, shapes the way EAP teachers view the emphasis placed on language use, and arguably the appropriateness of particular language use. When asked which instances of language use may lead to lower or higher marks in the AOP assessment on their specific courses (in question 3.4 in Appendix B), the themes which emerged in the data were: errors in text on slides (7 practitioners); grammar (5 practitioners); register (2 practitioners); verbal referencing (1 practitioner); vocabulary (6 practitioners); hesitation/pace (3 practitioners); intonation (2 practitioners); pronunciation (5 practitioners); memorisation or reading affecting oral delivery (5 practitioners); and language use which impeded communication (4 practitioners) or affected intelligibility and engagement (3 practitioners).

Some lecturers focus on discrete language elements (see EP 5) while others prioritise wider issues of overall communication and engagement (see EP 11):

Word choice

Third person s - pronunciation

Subject-verb agreement

spelling mistakes on slides

EAPQ, EP 5

Those which impede communication and engagement

EAPQ, EP 11

EP 5's response suggests that this AOP teacher assessor focuses on specific linguistic features which arguably may not affect task completion -- this is a perspective shared by teachers and students in S2 (see Sections 6.7.3.2 and 7.4-7.5). This evidences that there is an argument to be made for exploring how language assessment practices match with the overall AOP task purposes and characteristics. A strong case for questioning the validity of the approach to assessing language surfaced in data and is discussed next.

4.4.5   Difficulty Noticing Language Use

The ability to treat language use is influenced by how language use is perceived and processed by raters, and therefore data pertaining to difficulties in noticing language use contributes to addressing S1 EAP module data sub-RQ4: 'How is language use treated in EAP AOP assessments?' In answer to question 3.5 pertaining to challenges in the EAPQ (see Appendix B), 3 EAP practitioners express difficulty 'noticing' language use when assessing the AOP and some questionnaire respondents including EP 13 (see comment below) explicitly stated that the non-linguistic criteria distract from marking the linguistic features. EP 6 believes that the criteria other than language were more easily noticed.

> All because there is so much to think of in one go.
>
> probably language (use of appropriate vocab and grammar) as easier to reflect back on content/argument and easy to notice delivery.
>
> EAPQ, EP 6,
>
> Trying to distinguish what consists of "poor use of grammar" - as it is quite hard to notice grammatical errors in speech (much easier when marking essays)
>
> EAPQ, EP 12
>
> It is sometimes hard to give feedback on individuals grammatical errors because there are other aspects including content and presentation skills, which require more focus to mark.
>
> EAPQ, EP 13

The difficulty noticing language is a topic which has surfaced when talking to colleagues during my time assessing EAP AOPs and has been mentioned by 3 lecturers out of 15 in an open-ended question about challenges, which indicates it is an important issue to consider. Even if mentioned in one case, this is an area which warrants further exploration. Not being able to 'notice' linguistic features in the criteria has implications for test validity. What is meant by 'noticing' is an area to be investigated. Also, whether raters wish to notice language to give feedback or to make assessment decisions would need to be ascertained when considering potential ramifications for the assessment validity. This issue is at the heart of why the assessment task purposes, functions and specific elements assessed and interaction between them is an area which is explored in this project. The need for the criteria to closely cohere with the AOP task characteristics and with what is feasible and logical for teacher assessors and students to achieve and prioritise is a key issue raised in S1 and S2 and is more fully discussed in 8.5 and 9.2.1.

4.4.6   EAP Module Analysis Summary

An examination of EAP module data in S1 has revealed a number of interesting findings, particularly a number of tensions. The data analysis has shown that there are diverse task formats but that language is a key criterion in EAP AOP assessments and the aspects assessed were more limited than in subject module courses. A discernible contrast in EAP practitioners' stated approaches emerged in the data analysis between AOPs used to assess linguistic skills in isolation and those assessing a broader integration of skills needed in preparing and delivering an AOP performance. This contrast represents a tension in what is viewed as the purpose of EAP and the EAP construct. Is the primary role of EAP practitioners to develop linguistic skills or a suite of skills necessary for academic study in English medium institutions? This tension surfaces in AOP rating practices in the treatment of linguistic and non-linguistic skills, according to EAPQ respondents. They recount experiences of struggling to isolate and assess linguistic performance while

maintaining an ability to digest the AOP performance as a whole. This important validity concern and social justice issue was identified as a vital area to explore when conducting S2 fieldwork, in this study on assessment practices. The next section summarises the key broad findings from S1 data and shares how these refined the leads to follow ahead of S2 fieldwork.

4.5   Chapter Summary

In this chapter, I have presented an analysis of S1 data and shared the findings of the first phase (out of two) in this project. The discussion of analysis of data and findings from S1 and S2 is contained in Chapter 8. The main findings from S1 are that AOP assessments in HE key into diverse skills and fulfil a myriad of functions. AOP delivery skills may be crucial to AOP performances and therefore oracy demands associated with AOPs warrant close investigation in both EAP and subject modules. In terms of EAP AOP stated practice, distinct approaches and potentially problematic tensions emerged from an analysis of EAP marking criteria and EAPQ responses. The AOP may be used predominantly to tap into linguistic skills or a blend of linguistic and non-linguistic skills. The test takers' ability to deliver an AOP as a communicative act may be assessed or capabilities regarding discrete skills may be gauged through the implementation of an AOP task. EAP practitioners may be tasked with extracting evidence of linguistic performance while capturing evidence on a range of non-linguistic skills; this act is presented as cognitively demanding for some EAP practitioners.

In terms of findings regarding the methodological approach, the questionnaire data served to aid progressive focussing and priming for the complexity of AOP assessment practice. With the opportunity to reflect on S1 methods after having conducted S2 ethnographically-oriented fieldwork, the S1 methods only begin to capture the intricate interwoveness of factors including the validity and ethical quandaries teacher assessors navigate in AOP assessment practice. The findings outlined in this chapter are ones which required fleshing out further. Given the feasible scope of this project, I decided to focus attention on investigating EAP AOP practice, but further research on subject module AOPs is welcomed. I chose to go forward to S2 with central wishes and concerns, which were subject to change throughout S2. These concerns and wishes were to focus on teachers' and students' perspectives, investigate practitioner and student actions as well as perceptions and examine the myriad of factors which shape practice.

I devised the following research questions during and after fieldwork. As S2 fieldwork and analysis progressed, access requirements and the focus evolved. The questions presented below were finalised in the closing stages of this project, but reflect the approach and spirit in which I

approached S2 fieldwork throughout. The main purpose of their inclusion in this thesis is to serve as useful reference points in digesting the following chapters.

Guiding sub-research questions S2:

1) What are the teacher and student conceptions of the EAP AOP assessment purposes on their course?

2) What do teacher assessors do in EAP AOP assessment practice on their course?

3) What factors shape teacher assessor AOP assessment practice on their course?

4) What factors shape student perspectives on EAP AOP assessments on their course?

During fieldwork, due to apprehensions as to managing data and asking too much of informants, I observed and analysed teacher actions more than student actions. I observed students in classes and their AOP performances in real-time. However, in the research questions I state that I explored student 'perspectives' in contrast to examining teacher assessor 'practice' and what teacher assessors 'do'. Had I observed students preparing AOPs and analysed the video-recordings of AOPs, I would have opted for investigating student 'practice' rather than 'perspectives' in this research output.

The next three chapters present the analysis of The Fieldwork Study data.

Chapter 5   Teacher Assessor Processes in AOP Assessment

5.1   Introduction

Chapter 4 reported on data analysis of S1 data which provides a crucial backdrop to S2. The manifestations of AOP tasks reflected in documents and teacher accounts of their practice in S1 pointed to crucial phenomena to explore (e.g. the difficulty noticing language use) and gave useful diverse reference points which could be used to better understand local practices at the field sites in S2.

The next three chapters present S2 data from fieldwork conducted at two sites - two EAP modules at two different universities in the UK. The following table gives the achieved data set in S2 divided by field site:

Table 12: S2 Achieved Data Set

Source	FS1 Data	FS2 Data
Documents	Module year-round syllabus Rating scale for AOP task Rating scale from last year's AOP task	Module description Rating scale for speaking tasks
Teacher interviews *Captured through audio-recording*	2 interviews with 1 teacher Member checks with 2 teacher assessors	2 interviews with 1 teacher Member checks with 2 teacher assessors
Student interviews *Captured through audio-recording*	2 interviews with 4 student informants 1 interview with 1 student informant	2 interviews with 3 student informants
Rating discussions between teacher assessors *Captured through audio-recording and field notes*	1 rating discussion (allotting marks for 14 students)	1 rating discussion (2 sittings – allotting marks for 11 students)
AOP student performances *Captured through video-recording*	4 Group presentations (14 student presenters in total)	11 Individual presentations (11 student presenters in total)
Observations of lessons *Captured through field notes*	Observation of 3 lessons	Observation of 3 lessons
Field notes	Field notes taken throughout the fieldwork	Field notes taken throughout the fieldwork

This chapter and Chapter 6 focus on what the teacher assessors do most closely related to the AOP assessment. This chapter documents the processes teachers engage in when designing and implementing AOP assessment. Chapter 6 offers a close analysis of what factors into decision making when rating AOPs. Chapter 7 gives a detailed analysis of factors which shape how students approach AOP assessment tasks. The analysis in Chapters 5-7 will show it is not only what happens during the AOP preparation, the AOP assessment events and the AOP rating discussion which contribute to AOP decision making. Learning and assessment activities on the modules are inextricably tied together and multiple interactions outside the core AOP-centric tasks shape AOP assessment decision making.

In analysing S2 data, I lay out processes that the teachers are involved in when designing and implementing AOP assessments and what influences these processes. Studies which have investigated processes in classroom assessment (Rea-Dickins, 2001; McNamara and Hill, 2011) have produced frameworks of classroom assessment processes which tend to have a chronological structure to them: for example 'planning', 'framing', 'conducting' and 'reporting stages' (Rea-Dickins, 2001). The processes set out in this chapter may seem, at first glance, to have a chronological logic, but these are ongoing activities which vary in their level of intensity throughout the lifespan of teaching and assessment processes, on and beyond the module. The core categories have arisen from analysis of documents, interview and rating discussion recordings and transcripts and field notes made throughout my time in both sites. Data from interviews before AOP assessments (hereafter IBP) and interviews after presentations (henceforth IAP) are included. The core processes are: Inheriting and Influencing AOP Assessment, Constructing AOP Task Representations, Communicating AOP Task Representations, Gathering and Selecting Assessment Evidence and Critically Reflecting on AOP Assessment Practice (discussed below in Sections 5.3-5.7). It is important to investigate such core actions when endeavouring to develop a closer understanding of validity in AOP assessment. Having a grasp of these overarching substantive processes (and their dimensions, properties, relationships and variations contained within them) in the assessment design and implementation provides a vital means by which to better understand score-reaching decisions (the focus of Chapter 6). The analysis here points to what shapes decision making, addressing RQ2: 'What factors determine the EAP AOP assessment construct?' Before I discuss the core teacher processes in AOP assessment, key details of the AOP assessment tasks are outlined.

5.2   **The AOP Task**

This section provides the key details of the task format and the rubrics in the two field sites. In Section 5.2.1, I present (in Table 14) the task information relayed by teacher assessors to students orally or in written form, and I also give a description of the core details of the AOP task from my experiences and observations in the field sites (in Table 15). I devote Section 5.2.2 to an analysis of the rubrics (also referred to as rating scales) used in both sites.

5.2.1   **The AOP Task Format**

The AOP task details were communicated in documents and in class. This information is included in Table 14 below:

Table 13: AOP Core Details

	FS1	FS2
Presentation type	Group	Individual
Assessment task details	Group presentation which shares findings of a group research project. This group research project tasked students with designing a questionnaire for home and international students on a topic of their choice and then obtaining responses.	Individual presentation on a topic of student's choice (agreed by the teacher).
Timing	7mins per person in the group + 5 mins Q&A for whole group	10mins presentation and 5mins Q&A

The assessed presentation in FS1 required students to undertake a group project which involved research skills and group work skills. In FS2, the presentation involved intensive research on a topic. The timing of the assessments stated in task briefs did not necessarily happen in practice. In fact the question and answer phase was not exploited in many of the AOP performances.

From observing practice in the field, extended description of the AOP events is necessary to understand the complex processes which will be presented later in this chapter and in chapters to follow. I considered format characteristics which surfaced in S1 survey analysis which included categories such as 'audience' (as outlined earlier in 4.2). The next table gives core AOP details according to my observations in the field which were not included in communications of assessment tasks I accessed in the sites.

Table 14: AOP Description for Research

Audience	Live audience of 2 teacher assessors and students The presentations were video recorded The recording was watched by teacher assessors after the event to facilitate score decision making and used for external examination purposes	Live audience of 2 teacher assessors and students The presentations were video recorded The recording was available for moderation and external examination purposes
Assessed by	Two teacher assessors who have taught the students prior to the assessment	Two teacher assessors who have taught the students prior to the assessment
Assessment type	The presentation takes place at the end of semester one on a whole-year course and is the first of two presentations; one of which will contribute to a portfolio. It is weighted at 10%.	The presentation takes place at the end of a whole-year course and is the one 'speaking assessment' which contributes to students' final mark. It is weighted at 25%.

The length of the course, the number of assessors, the use of video recording, and the live delivery of the performances are all crucial features of the task needed to understand decision making. The rubric also becomes relevant to understanding the choices taken by teacher assessors and students. In particular, the rubrics become a useful analysis tool in gauging significant agreement or departure with the predefined conceptualisations of the AOP assessment tasks.

5.2.2 The Rating Scales

Both field sites use a rating scale -- also referred to as a rubric and marking criteria -- which consists of a matrix with criteria. The criteria come in the form of prose descriptors (e.g. 'use of grammar significantly affects intelligibility'). These descriptors are separated into sections (e.g. 'language'). These sections fall within particular bands of marks (e.g. 0-20). As outlined in Section 3.10, the full rating scales have not been published here; the names of sections have also been slightly changed so as to minimise the potential for the identification of institutions and, consequently, of key informants.

In FS1, the rubric consists of five sections all equally weighted at 20% of the mark. There are five bands which each contain 20 marks; the assessment is out of 100. The FS1 rating scale sections are provided in Table 16 below.

Table 15: FS1 Rating Scale Sections

Sections
Content (20%)
Fluency and Intelligibility (20%)
Accuracy and range (20%)
Presentation skills (20%)
Group work (20%)

An analysis of the rating scale indicates that the AOP assessment, assuming the AOP performances are mapped to the scales, is an integrated skills assessment with a greater focus on linguistic skills than on other skills (group work, content and presentation skills) in the assessment. Linguistic performance accounts for 40% of the grade overall though signalling that the rating scale fits mainly into the 'AOP as a communicative performance' category (outlined in the typology in Table 12 in Section 4.4.2). I would class this rating scale as a strong sense of performance assessment (McNamara, 1996). McNamara describes a weak performance assessment as having assessment criteria heavily linguistic in focus. I interpret a scale with the linguistic criteria accounting for 50% or less of the weighting as a strong performance assessment. The true emphasis on linguistic and non-linguistic skills, and level of adherence to the rubric, are areas to be gauged from analysing teacher assessor thinking and actions (unearthed in the remainder of this chapter and Chapter 6).

In FS2, the rating scale comprises six sections. There are six main bands with a mark from 35%-100%. The main rubric bands include reference to the Common European Framework (CEFR) levels from level B1 to level C1+. The descriptors for 0%-34% are less comprehensive and summarised on a separate sheet to the main rubric. This lower band references A2+ level or less in the CEFR. This sheet dedicated to low fails covers scenarios such as academic misconduct. The weighting for sections was unclear to teachers in FS2 – Georgina remained uncertain as to the exact weighting but assumed equal weighting for each section. The table below provides a summary of the FS2 rating scale sections.

Table 16: FS2 Rating Scale Sections

Section
Overall competence in academic English
Content
Fluency and coherence
Lexis
Grammar
Pronunciation

The FS2 rubric consists of academic skills and linguistic skills criteria. This clearly aligns with the 'linguistic and academic skills' category (provided in the typology in Table 12 in Section 4.4.2). It is difficult to confidently ascertain the extent to which language is emphasized in this assessment due to ambiguity as to the weighting for each section, as discussed above. However, the descriptors in the 'overall' category are framed by students' linguistic competence indicating the assessment is viewed primarily through a linguistic lens. Furthermore, the language categories ('fluency and coherence', 'lexis', 'grammar', and 'pronunciation'), in terms of the formatting, are presented as being separate sections on the same level as content. Therefore, I regard this rubric as indicating that the assessment would be a weak sense of performance assessment (McNamara, 1996). In other words, the criteria are heavily linguistic in focus. The teacher assessors may or may not show a strong inclination to apply these criteria to the assessment process. How teacher assessors' core actions and score-reaching decision making aligns or departs from the rating scales has ramifications for how the validity is enhanced or compromised in the assessments. Assessments which may be considered as weak senses of performance from analysis of the criteria may be in fact be re-defined as strong performance assessments when rater behaviour is captured and analysed.

How the rating scales were constructed is discussed in 5.4 and in the next section on influencing and inheriting the AOP task.

5.3   Inheriting and Influencing AOP Assessment

The grounded analysis, using CGT methods, of S2 data related to teacher assessor actions generated key theoretical codes. The core codes were developed through coding and memo writing in NVivo (See Appendices K and L) and further refinement in written drafts. One of the core codes outlined in this chapter is that of 'Inheriting and Influencing'. The properties of

'Inheriting and Influencing AOP Assessment' are 'Developing a Vision', 'Confronting Constraints on Delivering Vision' and 'Tweaking or Overhauling the AOP Task' (depicted in Figure 7 below).
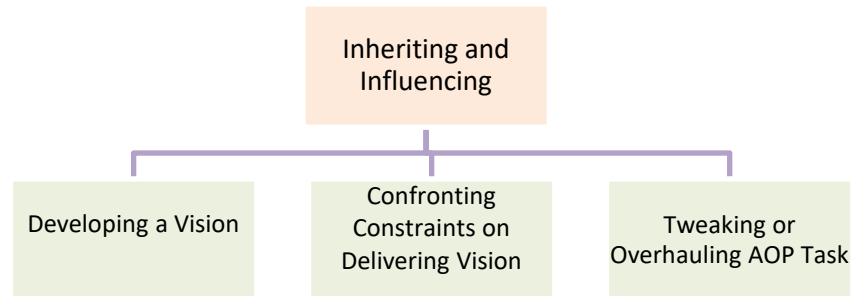
Figure 7: 'Inheriting and Influencing' Theoretical Code and Properties.

The remainder of section 5.3 outlines properties of this code through the inclusion of verbatim data and description which relates to the level of influence teacher assessors Adam and Georgina exert on course and assessment design at the two fieldwork sites. The extent to which teachers can develop their vision of the AOP assessment task is shaped by what artefacts and communications they inherit and their level of freedom and confidence to tweak or adapt the assessment task. The visions are abstract conceptualisations of what the AOP assessment involves and in some ways may be idealistic and removed from everyday practice at times. This is compared to an AOP task representation which is closer to practice, which is discussed in Section 5.4.

5.3.1   Adam's Influence

In our first meeting, Adam, the main teacher in FS1 (see Section 3.7 for setting and informant information), spoke enthusiastically about how he, after discussion with colleagues, had created the IFP module. He indicated his level of influence on the module design when he referred to it as 'his baby'. The module centered on encouraging students to 'examine their learning environment' and reflect on 'their state of being a student entering university and becoming a postgraduate student in the UK' (FS1 field notes). Adam used 'loose themes' of group work and reflection as course building blocks. Adam was inspired by Van Lier's (2004) ecological perspective on language learning which posits that 'all elements within the ecosystem -- such as a classroom -- are interrelated'(2010, p.5) and that there must be space for a variety of expressions of agency to flourish. Adam also recounted how he had wished to develop an assessment task which brought the group together and involved the students in the whole research process (see potential impact on student perspectives on the AOP task in Chapter 7). He believed a group AOP would best suit these aims. Group work is included as a section in the rating criteria because, from what Adam

has gleaned from working in a Higher Education institution, group assessment is a growing feature of UK HE.

The assessment design, according to Adam, is a reaction against Pre-sessional presentations which are 'just sort of shoe-horned into the end of the course' (FS1, IBP). Adam acknowledges the practical constraints faced on short Pre-sessional courses with high student numbers, but believes the way these are implemented discourages participation and engagement in the social context of university life and learning. The AOP, for Adam, has an overall purpose of developing student confidence and to promote student engagement in the social context of university. Adam had a great degree of input into the course design and has had a degree of influence on the existing rating scale, having made amendments to it at the beginning of the academic year (these amendments are discussed in Section 5.4 below).

This analysis demonstrates that Adam's vision was developed through dialogue with colleagues, knowledge of HE practices, exposure to academic thinking on pedagogy, reaction against practices on other courses, the freedom to develop the module design in his role as module leader and his general investment in the shape the module took. These conditions all act as facilitators to implementing his vision but constraints are also at work (see 5.3.3 below). Georgina's context in FS2 provides different conditions and changes the degree and type of influence which is exerted on the AOP task.

5.3.2   Georgina's Influence

Georgina, the main teacher in FS2, had been making changes to the EAP module on the IFP gradually over six years and she was highly conscious of ensuring a reasonable amount was expected of the students.

Georgina was keen to highlight that the learners require 'a fair amount of scaffolding', which she worked into the lessons over the years, in order to help prepare her students for their prospective undergraduate programmes. From my observations, I saw that Georgina gave feedback on students' AOP draft scripts and multiple practice performances, as well as providing intensive support with finding sources. She placed a great deal of emphasis on uptake of feedback and improvement because the students have the opportunity to deliver work on the AOP for a whole semester before the summative event. Georgina described her role as providing a 'safe space' for students where she exposes them to types of situations they may encounter on their prospective courses. She offers students opportunities to gain useful tools and strategies to cope on their undergraduate course.

Georgina required students to give an individual AOP on a topic of their choice and has started creating banks of resources on set topics. She invested a huge amount of time in ensuring the students, who would be going onto undergraduate programmes, were not expected to display a huge amount of complex skills without having had adequate input and support. Georgina's influence was exerted at the level of classroom practice. The level of study of the students greatly shaped Georgina's pedagogical approach.

Both Adam and Georgina recounted how they had shaped assessment design and implementation to a minimal degree. This modest contribution to assessment practice is the result of challenges Georgina and Adam face when attempting to establish and enact their visions of practice.

5.3.3   Barriers to Influencing

These teachers work in different institutions, and prepare students for a different level of study, but both Adam and Georgina communicated, in interviews and informal conversations with me, high levels of investment and influence over the course design and implementation. However, when talk shifted to assessment decisions, and the rubric in particular, both Adam and Georgina both communicated how they are not entirely comfortable with the assessment dimension of their practice: the discomfort was particularly pronounced in comparison with the confidence with which they spoke about pedagogical decisions. Adam talked about having 'inherited the criteria' and how he managed only to make some alterations to it. Georgina shared how she was involved with a group of teachers in designing the rating scale six years previously, and that when she arrived at the institution they 'were preparing a validation and nobody actually knew what they were doing' (FS2, IBP with Georgina). This pointed to a potential lack of assessment literacy resulting in teacher assessors not possessing the confidence to devise and implement assessments.

Georgina also recounted how she wished to introduce a seminar style assessment but was not sure how to approach this. When referring to her influence on assessment practice, she spoke of making adjustments to the assessment rather than implementing a radical overhaul.

> Georgina: I think it's just unfair to ask them to do a presentation when you don't actually teach them the content. Because you know they're learning so but then how else do you assess language is the issue. Through a discussion conversation, maybe that's possible but I wouldn't know how to go about it. And I don't have the time to go changing these assignments. So all I want to do is adjust it so it's you know a bit more friendly. I mean they're COPING ok you saw!
> (FS1, IBP with Georgina)

The teachers have a vision as to how to take the courses forward, but how to create a valid assessment, particularly with a rating scale involved, is an area about which they expressed a

great deal of uncertainty; it was presented as somewhat of a stumbling block. The level of influence over course design was substantial, but in terms of assessment these teachers made gradual adjustments and admitted this is due to influences of standardized assessments such as IELTS, university policy, the introduction of tuition fees in universities in England, and their assessment literacy. In the extract below, Georgina recounts how the rating scales are based in part on IELTs and university marking criteria.

> Georgina: So this needs changing. Because this bit here is based on the IELTS marking criteria.
>
> Interviewer: The language part?
>
> Georgina: Yep, and the content is based on the university generic marking criteria. . . . You can change some things {in the module} like it might be you can't run the course or whatever. But once the course has started that's it. You publish the stuff {programme documents including criteria} on day one and if you don't like it tough because students have paid £9,000. Right so you can't complain.
>
> Interviewer: So for EAP and content courses that's the same.
>
> Georgina: Any courses at the university or any university, as a customer, a paying customer as students, they give you a piece of paper and that's it on day one and no changes.
>
> (FS2, IBP with Georgina)

The excerpt above highlights influences on rubric design. Moves for consistency in rubrics across institutions mean EAP rubrics may be mapped to university criteria. IELTS, as an assessment which is seen to have gone through a validation process, is seen as a trustworthy reference point, even if classroom assessment differs greatly from large-scale standardized assessment. In FS2, Georgina received training which actively discourages changes to assessment and module documents. Georgina links this practice to the marketization of HE and measure to safeguard against complaints from students, or 'paying customers'. While publishing assessment details and criteria at the beginning of the course ensures a degree of fairness, this leaves little space for negotiation of assessment expectations. This is shown to be particularly challenging for teachers on courses when there is prolonged and direct contact and interaction with students. Adam and Georgina talked about the organic nature of teaching as it is shaped by the teachers and students on the module throughout the course of an academic year. The constraints mentioned limit possibilities and result in tweaking practice instead of overhauling it. The teachers have a vision as to what is possible in their practice, but when implementing tasks they develop pragmatically oriented conceptions as to the AOP task. The next section reveals how task representations are not fixed and determined wholly by rubrics and task briefs set at the beginning of a module. The visions of the task can be shaped by evolving and emergent practice.

5.4   Constructing AOP Task Representations

The teachers, throughout my time with them in the field, divulged what they understood to be the skills, abilities and processes required in the AOP assessment - their *task representations*. Flower et al. (1990) define task representation as 'an interpretive process that translates the rhetorical situation—as the writer reads it—into the act of composing' (p. 35). This definition relates to written tasks but can be applied to AOP tasks and not only to the person creating the artefact (presenter, in our case the student test taker) but those judging the piece of discourse (the audience, in our case the teacher assessor, students and researcher). The analysis of S2 data demonstrated that the teacher assessor representations are developed and clear in many respects, but ultimately fluid constructions. Teachers inherit and influence documents and policies building a vision of the task which is open to change. The visions of the AOP task, discussed in the previous section, which are developed from engaging in the core process of Inheriting and Influencing are broader abstract conceptualisations operating at a higher conceptual level than the AOP task representations. The representations are less abstract notions which result from implementing the assessment and taking action. Engaging in such action with close proximity to the realities of practice has the potential to change the visions of practice and the intensity and level of influence. The proximity to practice, for instance, may reveal tensions between official guidelines and unofficial practice. Figure 8 below represents the theoretical code 'Constructing AOP task representations' and its core properties:
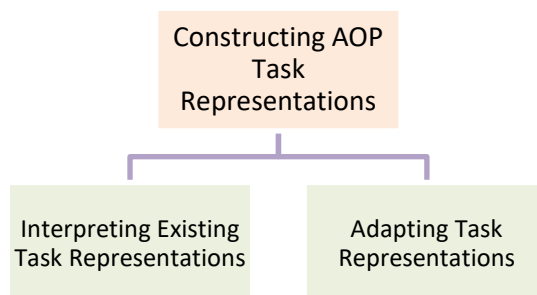
Figure 8: 'Constructing AOP Task Representations' Theoretical Code and Properties

In my time in the field, I detected shifts in the teacher assessor interpretations of the tasks. In this section, to evidence this finding in the limited space available, I have chosen to present the analysis of how teachers' reflections clearly evolved on the role of notes and non-verbal communication. Their evolving stances on the use of notes evidence changing perspectives on the AOP task and what they expect from student test takers. I begin by sharing changes to Adam's unique representation of the AOP task before sharing how Georgina's task representation evolved.

5.4.1   Adam's Task Representation

At FS1, Adam made changes to the rubric at the start of the semester. An analysis of the rubrics before and after these changes, reflect, in part, how Adam created his unique representation of the AOP test. The changes included minor amendments such as removing 'volume' and adding wording to improve the consistency across the bands in the language and presentation skills sections, as well as adding 'cohesion' to the teamwork section, among others.

 In addition, Adam added specific descriptors in the lower bands of the presentation skills section in the criteria regarding the role of memorisation and notes: 'memorised chunks' and 'overreliant on notes'. The inclusion of memorisation and the intensification of specific descriptors in reference to script dependence in the lower bands sends a signal that students should move away from use of scripts and deliver in a way that does not appear memorised. Adam talks about how he made these changes in order to shift students away from mechanical delivery:

> Adam: I think what I'm trying to suggest with overreliance on notes is also that it's sort of damaging the presentation because they're not making eye contact. So I can't really see how if you were sort of practically reading notes of memorising bits that you're actually giving a proper live presentation to a live audience. The presentation should also be interactive in a sense… So it shouldn't just be like kind of one way memorised language. (FS1, IBP with Adam)

Adam here is emphasising the requirement to speak in a way that seems less mechanical to improve audience engagement. In interviews, Adam mentioned how he is 'quite a stylist' when it comes to presentations because he gave presentations at schools when he worked abroad. Adam's past experience related to presentations is one factor which determines what he deems to be a successful AOP performance and prompts him to create a new task representation from the AOP task in his module last academic year. He has processed the existing interpretation and created a unique new interpretation which has evolved gradually over time.

5.4.2   Georgina's Task Representation

In FS2, Georgina did not make changes to the rating scales during my time with her. In FS2, I traced a slight shift and ambiguity in Georgina's views regarding script reliance, however. In sessions on practice presentations, Georgina advised students to 'drop the reading'. Then in discussions (prior and after the assessment) she considered how realistic and fair it would be to expect script independence. She cited examples of home students in the target domain using notes in presentations. She also considered how feasible it is for the students to internalise the content of their presentations.

Georgina's first statement, to her co-assessor Tracey, just after having assigned summative marks for half the group was: 'Do you think they should be learning it by heart?' (FS2 rating discussion). Georgina and Tracey then entered into an exchange considering what is fair to expect of their students. They ultimately posed the question of whether being a good presenter would be necessary to their students' survival on an undergraduate programme, but the question remained mainly unanswered. Georgina paused the conversation and the three of us head to the cafeteria before the second session of summative presentations. Tracey and Georgina picked up this conversation again in the second rating discussion after lunch when talking about a student's AOP performance:

> Georgina: Reading throughout.
>
> Tracey: Yeah
>
> Georgina: So that I don't know I'm never sure about this. I've seen domestic students reading
>
> Tracey: Yeah.
>
> Georgina: To me if you start reading that's no good. But then you go to a conference, like remember {Name redacted}.
>
> Tracey: Yeah.
>
> Georgina: And that woman who read her whole paper.
>
> Tracey: Yeah and I've been to the same.
>
> Georgina: NO BUT SERiously so how can you then tell a STUdent not to read right?!
>
> Tracey: It's similar to what we said this morning isn't it. I've put {in real-time notes} shame you read your notes, it affects your pronunciation, you lose your sparkle. You've chosen a difficult topic to talk about and it's a shame that you didn't practise more
>
> Georgina: Yeah.   (FS2, Rating Discussion Part two)

This exchange during part two of the rating discussion in FS2 shows the range of issues tied up with reading in the presentation. Reading may affect audience engagement, intelligibility, and may be perceived as lack of preparation. On the other hand, use of notes and a script is not uncommon in conferences and on degree courses and is not necessarily viewed unfavourably. In many fields presentations involve reading a prepared paper.

 A week after the presentation, Georgina and I discuss this quandary in our interview after the presentation, and Georgina at this point is considering the ethical dimension of assessing presentations skills: she now considers these to be 'bullshit skills' that people in many spheres of life put too much emphasis on. Georgina believes the content should be the main priority. Georgina wishes to develop her students' confidence to speak freely and encourages students not

to read, but does not feel it is fair to deduct marks for script dependence and assess presentation skills such as eye contact, body language, and position. Georgina's view on use of notes and the importance of presentations skills has evolved over the duration of the course.

Adam and Georgina's contrasting views and changes to their cognitive models of the task -- relating to script dependence and importance of presentation skills -- supports the assertion of Flowerdew et al. (1990) that task representations are created rather than selected. This data shows that the assessors, as well as students, also create task representation over a period of time. How teacher assessor representations translate into a coherent, workable and valid assessment construct is the core conundrum (as discussed in 2.10.2 and taken up in 8.6). And how evolving task representations are negotiated with students is a key issue at stake (see Chapter 7 for information on student task representations). The teachers are responsible for *communicating* their expectations (which has been shown may be changing and in a stable or ambiguous state) in a clear and consistent way.

5.5 Communicating AOP Task Representations

Teachers communicate particular task expectations to students prior to, and during the assessment itself. The communication of task information as well as taking place at various stages in the assessment events, also comes in the form of explicit and implicit guidance. Explicit guidance refers to instances when the teacher assessors very clearly state what they expect in documentation, verbal and non-verbal communication (e.g. shaking head at student, 'I wouldn't recommend using a script'). Implicit guidance refers to cases when expectations may be communicated in a less direct way and there is more potential for students not to pick up on such cues or nudges. As will be demonstrated in the ensuing analysis, teacher assessors gauge students' interpretations of the AOP task, and provide explicit guidance and nudges to improve learning and bring the students' interpretations closer to their own. The core properties of the central process 'Communicating AOP Task Representations', which were developed through coding, memo writing and drafts (See Appendices K and L), are depicted in Figure 9:
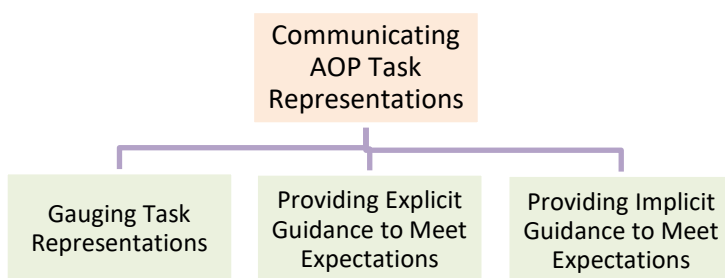
Figure 9: 'Communicating the AOP Task' Theoretical Code and Properties

5.5.1   Explicit Guidance

The teachers at both sites provided task details and the rubric to students in advance of the formal AOP assessed performance (see Section 5.2.2 above for summaries of the rating scales criteria). These provide explicit guidance as to what is deemed important, even if there is room for interpretation. The prose descriptors and numerical bands send a message as to what will count towards their mark. However, the teacher assessors may give conflicting advice or indicate flexibility in these guidelines.

In lesson observations, in both FS1 and FS2, the teachers devoted class time (one lesson in FS1 and two lessons in FS2) to practice presentation sessions in which they gauge students' task representations which is shorthand for determining their level of understanding of what the teacher assessors deem to be required. In these sessions, teachers and peers give explicit guidance to the presenter on content, structure and timing, verbal and written language use, use of slides, preparation, body language, pace, referencing, energy and enthusiasm and use of notes (themes written in FS1 and FS2 field notes). An example of explicit guidance is: 'Overall, I think you're there. Drop the reading though' (FS2 field notes). The criteria in FS2 contain no mention of reading from notes and the students, therefore are confronted with navigating another expectation in the explicit guidance. Adding to the complexity of potential mixed-messaging – or nuances -- in explicit requirements and feedback is the implicit guidance delivered by teacher assessors in nudging students to meet expectations.

5.5.2   Implicit Guidance

There were episodes in lessons in which the assessment expectations were made explicit but also times when expectations were communicated in the form of gentle nudges. Before the assessed presentations, Adam delivered a lesson on describing graphs and provided students with a variety of ways to present their data visually. He gives the group of students the rationale behind the activity: 'I thought it would be similar to what you do in your presentation' (FS1 field notes). This is an explicit signal to incorporate what you learn in this class into AOP assessments. In the implementation of activities, there is a layer of explicit advice. Adam gives the students an opportunity to stand at the front of the class and present their graph. At the end of the lesson he gives the students some pointers on language use, handling questions and position. Adam told his students 'Don't forget your audience is over here' and 'if you get asked a question, don't just say yes or no' (FS1 field notes).

Entwined with this explicit guidance, were threads of nudging at work. Adam revealed in an interview that he had numerous objectives in mind in this lesson, including 'ulterior motives' that were not communicated explicitly in the lesson:

> Adam: The sort of ulterior motive is saying these are the different types of graphs that you could use hehe. And look you could use a vaRIEty if you wanted to! So that's the ulterior sort of stealth motive that's in there as well. But then there's also the main one which was just to get them speaking about graphs and how to describe them. (FS1, IBP with Adam)

Whether students pick up on the cue to use a range of graphs emanating from Adam is not clear. As well as having what Adam calls his 'stealth motives' in the classroom, Adam brings these into the assessment event. In my field notes, during the AOP assessment I noted down how Adam was not looking at the students for the first part of their presentations. In an interview I asked him to talk me through the rating process and he revealed what prompted him to look down:

> Adam: I'm very conscious that when you're the teacher that students who are presenting will stare straight at you because they want to present to you. So I quite often start by just looking down and just listening for the first minute or so because I think that kind of encourages students not to just look at you the whole time which tends to happen a lot. (FS1, IAP with Adam)

Adam's strategy is aimed at nudging students to scan the room in order to improve audience engagement. This was the only instance of nudging I detected during the assessment process itself in the two sites. This has been categorised clearly as implicit guidance as Fiona (a student) and I were not aware of the reason behind this. I remember thinking – 'Adam look up, you're missing the slides and body language!' I identified with Adam's strategy because I have developed my own eye-contact nudging technique! When students fix their pupils in my direction, I scan the whole room in an overt and contrived manner, hoping they will pick up on this cue and divert their gaze. There is an impulse for teachers to want the students to do their very best and as an assessor it is an act of great self-control to remain nonchalant, especially when your in-built instinct is to help – arguably common amongst teachers. A part of real-time marking is supressing your reactions and considering how teacher assessor facial expressions and body language may affect student performance (see Section 7.6.2 on students' awareness of this dimension of AOP performance). This is a practice which resonates with me and a telling episode as it reveals that nudging students to meet expectations may not only happen prior to the event, but during the formal assessment event itself.

In these educational assessments, teachers explicitly give guidance and implicitly nudge students to meet expectations shaped by their own interpretations of the task. Teachers and students require an awareness of such features of classroom interaction. In such cases, fairness can in fact be strengthened by insisting teachers mark their own students as they are aware of the explicit

and implicit guidance given. Caution is to be exercised in gauging what guidance teacher assessors can expect students to have detected and act upon. It is also crucial that teacher assessors remain alert to potential influences of nudging behaviour during the assessment. An awareness of the expectations communicated by teacher assessors is needed. The explicit and implicit guidance may need to be declared during rating discussions to ensure fair expectations are placed on students. This is particularly vital in classroom-based educational formal AOP performance assessments (see discussion of how reference to guidance given is used in reaching score decisions in Chapter 6 and a discussion on enhancing validity in AOPs in Section 8.6).

5.6   Gathering and Selecting Assessment Evidence

The teacher assessors locate evidence and theorise as to a students' abilities within the AOP performance and outside of it. The AOP performance denotes the time when students deliver their presentation in front of the teacher assessors and peers – in other words, the formal assessment event. Evidence external to the AOP performance may include performances in other assessments, performances in the lessons on the module linked to this assessment and other modules on the programme. Inextricably tied up in this process of assigning a mark is the act of discerning the relevance of pieces of evidence. The teachers decide whether to select (to admit or discard) the proposed pieces of evidence. This process of selection is touched upon in this section (5.6), but is elaborated on further in the 'Critically Reflecting on AOP Practice' section (below in 5.7).
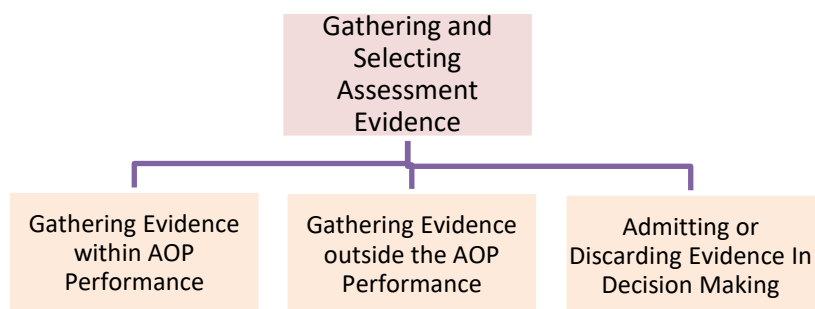
Figure 10: 'Gathering and Selecting Assessment Evidence' Theoretical Code and Properties

The code and properties are evidenced and fleshed out in the resultant discussion in sections devoted to references to data in which teacher assessors focus attention on AOP performances (5.6.1) and episodes in which teacher assessors venture away from the assessed AOP (5.6.2).

5.6.1   Within the AOP Performance

The teacher assessors use their real-time notes on the AOP assessment, videos of the AOP (FS1 only) and their embodied experiences and memories of the AOP events when reaching score decisions.

The notes made during the performance are heavily drawn on as a basis to formulate interpretations of what happened in the AOP assessments. However, the validity of the comments made in isolated moments during the presentation are questioned by teacher assessors' holistic interpretations of the events (see an instance of this in Section 6.3). From observing the teachers during the AOP performances, I noted that they 'were writing notes at the end of the presentations and after' (FS2 field notes). The teachers shared in rating discussions and interviews that they had changed their interpretations as the AOP delivery progressed and often assigned tentative marks at the end of the presentation.

In FS1, an extra source of evidence was at hand during the rating discussions. Adam and Daniel use excerpts from video recordings of the assessment when transitioning to discuss a new group of presenters and, in some parts, to consider performance of individual students on specific features such as body language and accurate language use. In the following reference, Daniel and Adam check a student's body language.

> Daniel: She's talking without notes, she's got nice hand, Ok she's looking at the floor. Well
> Adam: No she's not! She's looking around quite well I think.
> Daniel: Yeah
> Adam: No 11 is OK.
> Daniel: Gosh, these {video recordings} are useful aren't they (FS1, rating discussion)

The videos are particularly useful in FS1 due to the fact there was a two-day gap between watching the presentations and assigning the marks. The use of real-time notes and memory of events was not always deemed sufficient and reliable by Daniel and Adam. The notes may not contain comments on every aspect deemed relevant by the assessors. The comments written at one point in the performance, as with the snippets of the video chosen by the assessors, may also not be a representative sample of the holistic performance on given aspects. The memories of the performance as a whole are crucial for assessing the representativeness of a judgement made at a given point in the performance (see Section 6.3 for a discussion of the zoom strategy used to maximise validity). The assessors put forward the evidence from numerous sources and decide whether to admit or discard these in making judgements on student performance.

The teacher assessors navigate pieces of information internal to the AOP performance from multiple sources. In addition to grappling with often problematic evidence, they also refer to information from interactions external to the formal assessed AOP performance.

5.6.2   Outside the AOP Performance

The assessors use their previous histories with students and understandings of practices in the target domain to inform judgements.

The assessors draw on their experience with the students in class and notice when a performance in a formal assessment does or does not match the judgements they have formed on the person in 'planned assessment opportunities' and 'incidental opportunities' in class. (Hill and McNamara, 2011). Daniel and Adam have both taught Anna prior to the assessment. During the rating discussion they describe Anna as a 'confident' student but are wary that she may be unaware that she is 'coasting' due to weaknesses in her academic skills. Here Daniel and Adam note that the presentation reflects a characteristic of Anna's work during the year – she is confident but coasting:

> Daniel: I thought she was kind of weak on content
>
> Adam: Yes.
>
> Daniel: But that's Anna. (FS1, Rating discussion)

In FS2, Georgina and Tracey also discuss characteristics of the AOP performance which they anticipate due to their perceptions of students' capabilities. Georgina often makes reference to the two practice presentations delivered prior to the formal AOP assessment in class as a springboard to consider the performances in the formal assessment.

> Georgina: Ok, for Marina in the practice {presentation} the main issue was the content. It was descriptive, not analytical. THAT err didn't change in the final presentation. No academic sources, she had {website name redacted} as her main source information   (FS2, Rating Discussion, part one)

Further analysis of the rating discussion reveals how the assessors in instances use assessment opportunities from previous interactions with students. Perceptions of the students generated over time and space provide raters with approximate coordinates to locate the test taker, acting as an anchor and starting point. The assessors can, and in instances do, draw up the anchor and lower it again in new waters when student AOP performances do not meet the anticipated level:

> Georgina: No
>
> Tracey: You know there wasn't that digging a bit deeper. I think that's what I'm missing. To be honest I was a bit disappointed with hers.

Georgina: Hmm.

Tracey: I did expect more.   (FS2, Rating Discussion, part 1)

Teacher assessors talk about their disappointment when the students do not perform to their expected potential in the AOP event. The teachers have in cases narrowed down the areas in the criteria in which they think the students will fall. Georgina in FS2 states before the AOP assessment that her student Marina will get a B+ for language, and Georgina and Tracey point out a student to me in the cafeteria before the assessments identifying the student as someone they believe will fail. The teacher assessors demonstrate that they shift their thinking if the performance does not match expectations. A question remains however as to whether there are times when assessors see what they expect to see. I detected episodes when teacher assessors acknowledge when the performance does not match their predictions, but I remain cautious to state to what extent teacher assessors tug at the anchors and take action to change the mark in line with the performance. Without accessing teacher assessors' internal monologues, it is only possible for me to gauge when the anchor is moved than when it is not.

In interviews and informal discussions assessors indicate that gathering and selecting evidence from the AOP performance itself is what would be seen as the most valid and fair procedure, and where they *should* concentrate their efforts. However, the assessors heavily refer to their previous knowledge of students and to other students' AOP performances in their score decision-making talk. What is it that prompts them to do this if it is perceived as something that should not be encouraged?

In my field notes during the rating discussion, I noted down how I was struck by Georgina's attention to detail that was 'at a level that seemed impossible for someone going on the real-time performances' (FS2 Field Notes). She shares what is behind her detailed and seemingly perceptive accounts of the AOP performances.

> Georgina: I mean now I already know who's going to get what more or less. Because there has been a lot of practice work {…} It would be difficult you know to just hear it once without having heard the student before. So I can know more or less what to expect. That is not always a good thing because that is going to impact the decisions but I already know, you know. I'm not going to, I don't know like Marina, I'm never going to look at the language bit beyond this page {in lower bands}. I already know where they are!
>
> (FS2, IAP with Georgina)

Georgina's admission here strongly suggests that the strategy adopted by markers to draw on their working knowledge of students points to an issue with the enormity and unmanageability of the task if listening to presenters for the first time. Using insider perceptions, generated from

evidence collected outside the AOP performance, is a coping mechanism for an assessment which has a high cognitive load.

Another factor which may go some way to explaining why assessors talk about information external to an individual's AOP performance is that they are considering the fairness of the assessment process. Georgina talks about the difficulty in assessing the AOP performance, when for an academic year they have been making decisions based on the person's interaction and behaviour in different settings and tasks (see further discussion on assessing the person in Section 6.7.2). I am uncomfortable with Georgina's conviction that she 'knows' how Marina and the other students in her module will perform. However, after spending time with Marina and Georgina, I can see how Georgina can build a strong case for her perceptions. Georgina has spent time with her students for a year in various teaching contexts. Georgina states that insider perceptions are 'not always a good thing' and it 'will impact the decisions' (FS2, IAP with Georgina) in the rating discussions. Georgina may not be recognising her use of previous interactions with students as a tentative anchor and this is why she is quick to share its flaws.

The use of information external to the AOP performance is used to act as an anchor and a consistency check (see Section 6.6 on The Ranking Filter) to increase fairness as reasonable adjustments may be made in deploying such a strategy (see more discussion in Section 6.7 on The Social Filter). However, there are questions as to how this can have the opposite effect by allowing biases to shape the marks. A closer examination of this issue would benefit teacher assessors because it seems that the working knowledge of the student is a crucial tool for them. Despite a huge trend towards anonymity in assessment, identities are brought into these discussions. The fairness lies in bringing these previous histories and identities in openly and explicitly telling students that this is happening and the rationale behind it. It strikes me that, as in many professions, an element of trust needs be given to teachers. Co-assessors, moderators, external examiners, assessment researchers may need to more closely consider *how* the assessors are using information external to AOP assessment performances, and whether it in fact strengthens the grounds for the ultimate high-stakes decisions that are being reached (see further analysis on this topic in Chapter 6 and discussion on validity in AOP assessment in 8.6). The teacher assessors consider whether to admit or discard the evidence put forward within and external to the AOP. This is most clearly shown in episodes when teachers reflect on their own assessment practices. Evidence of rejecting proposed evidence is discussed in the next section.

5.7   Critically Reflecting on AOP Practice

In my analysis of teacher assessor actions and thinking, I was struck by episodes in which the teacher assessors became acutely aware and critical of their practice. From coding the data, I identified a key process of 'Critically Reflecting on AOP Assessment Practice' which involves pausing parts of the decision making in which teacher assessors go through the motions of getting the assessment done. This pause is replaced with directing attention to evidence and justifications and the teachers display a meta-awareness of the act of assessing. The teacher assessors enter a state of critical reflection when they evaluate the quality of the rating techniques and decision making. This process is represented in Figure 11:
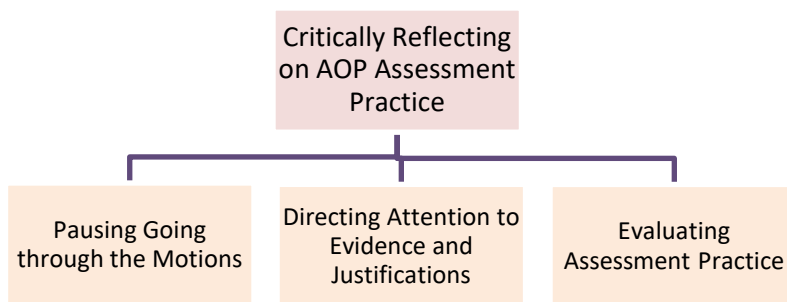
Figure 11: 'Critically Reflecting on AOP Assessment Practice' Theoretical Code and Properties

The remaining sections in this chapter demonstrate the inner workings of the category and its properties and what these look like in AOP assessment practice.

5.7.1   Going through the Motions

There are parts of the teaching and assessment process when Daniel, Adam, Georgina and Tracey appear to go through the motions. The code 'Going through the Motions' applied to instances, for example, when a mark is quickly agreed, with a reference to some key words in the rating scale and a negotiation of numbers. In some episodes, the speed and automaticity of the decision making was palpable with very minimal unpacking of how a score was reached. This is not to say that 'going through the motions' is to be viewed as a wholly negative phenomenon. In fact the 'automaticity' in some cases could be attributed to the virtue of the rating scale in making the task manageable teamed with a high degree of mutual understanding between assessors. Teacher assessors also refer to other student performances which are shorthand for (in some instances) earlier lengthy evaluations (more on this strategy in Section 6.6). The below example is an exchange in which Adam and Daniel agree one student's mark for fluency and intelligibility. They note a few key features, compare the student's performance to other students and assign a mark relatively quickly:

> Adam: What did you give him for fluency and intelligibility? I think he's very clear and easy-to-follow.
> Daniel: Yeah.
> Adam: Good pace and volume so I've gone as high as 12. Ok I don't mind. I think he's as good a level as Ivan or Henry or Anna
> Daniel: Yeah so 12 for fluency and intelligibility
> Adam: 12 for fluency and intelligibility. I think there's also a range there. So I don't know 11 or 12 for accuracy?
> Daniel: 11 I think.

From spending time with the practitioners, analysing their talk, and being present at the rating discussion, I experienced a strong sensation that the teacher assessors were thoughtful individuals who took great pains over the assessment decisions on the whole. However, they did not delve into the inner mechanics of the student test taker performances and their own rating strategies on many occasions. This seemed to be due to shared understandings of constructs and bands as well as due to time limitations. It could also be because of the dark matter and holes in their decision making – the teachers may not have ideas and evaluations to share on aspects of performances. Not debating the decisions in depth also may preserve a veneer of certainty which, in my own experience, seems to pervade much assessment practice, and to a certain extent is needed to keep many existing assessment systems operating. There are reasonable justifications for succinct evaluations in assessment decision making, however I am wary of teacher assessors slipping into such motions to too great a degree due to, for instance, time and resource constraints or their own disinterest and unwillingness to engage in the complexities and perplexities in the process. I am cautious of teacher assessors becoming automatons in the process which increases the chances for harm to be done.

In numerous other parts of the rating discussion, the assessment process is in a sense suspended (in both senses of the word) to reflect on the very nature of the assessment process itself. The assessors halt the task and hang it up to contemplate and dissect their actions: the complexities become more visible and tangible in the process. The enterprise becomes 'meta' and self-referential: the assessors assess how they assess the assessees.

5.7.2   Evaluating Assessment Practice

Pausing going through the motions occurs when teacher assessors slow the assessment decision making in order to focus more acutely at particular elements of a student performance. The teacher assessors tend to engage in detailed discussion and then closely evaluate their assessment processes when giving students low marks in both FS1 and FS2. In the following exchange at FS1, the assessors express a slight reluctance at giving Ben a low mark. They enter into a lengthy dialogue about Ben's fluency and intelligibility mark and then critically appraise the rating scales and the use of the bands within in it, before then determining which band Ben falls

within. The segment below is a small part of the prolonged dialogue relating to Ben's mark which illustrates how teacher assessors direct attention to evidence and justifications and evaluate their decision making behaviours.

Daniel: It's tough but we've got to put him in there. We've got to put him in there and I think.

Adam: Really? Ok.

Daniel: Well that chunk there I'm not sure if I understood it!

{…}

Daniel: I [think he's going to have]

Adam: [There's a big jump] between these two bands isn't there but in a way.

Daniel: Well it's also the way we're using them. We're giving most people the top of this band.

Adam: Hmm. but I don't think you can say he fits any of the criteria in the 8-11 band very accurately.

Daniel: No well that says it all then, doesn't it.   (FS1, Rating Discussion)

The assessors reflect on how the bands in the rubric may not have an incremental change in the level of the skills being tested, but how the way they use the criteria may help balance out any unevenness. Adam has shown some reluctance to give Ben a low mark but acknowledges that the Fluency and Intelligibility descriptors in the lower band fit his performance. This extract gives one example of how the teachers qua assessors evaluate their assessment practice, specifically how they work with the rating scale as human assessors: 'it's also the way we're using them'. The assessors consider the flaws in the criteria, but acknowledge its worth as a guide and accountability tool (see more discussion in Section 6.5 on The Criteria Filter).

Another example of teacher assessors critically reflecting on their decision making is when Adam and Daniel considered giving Tanya in a group presentation a higher mark for group work (20% of the AOP mark weighting) than the other students in her group. The teacher assessors discussed at length their interactions with Tanya prior to the performance. After the teachers considered the negative aspects of this student's performance, they decided she should receive the same mark as the other students. The student's efforts were not visible in the finished product. The teachers, therefore, ascertained that the 'insider knowledge' was clouding their judgement on this occasion because her effort did not translate into a successful performance. The following exchange is the latter part of their talk relating to Tanya's teamwork mark:

> Adam: My insider knowledge is I know that she did the presentation. But it's unfair because it's not from the presentation. But I think you can say from looking at the presentations, it's bad in terms of teamwork.
>
> Daniel: Yes, and I think that we're being distracted by your insider knowledge and in terms of the actual quality of the presentation she wasn't exceptional was she.
>
> Adam: No.
>
> Daniel: We know why. Yep we have the luxury of time. This is a group that we need to sit down and just say this one really didn't work because we could see the cracks.
>
> (FS1, Rating Discussion)

In this extract, Daniel critically evaluates the assessors' decision-making process itself when he states that the assessors are 'being distracted' by Adam's 'insider knowledge' that Tanya has done a lot more of the preparation work than her group members. They devote prolonged attention to the evidence they are using and the justification behind its contribution to the scoring decision. It is unclear whether Tanya would have received a higher teamwork mark if she had managed to hold the presentation together better though. However, on this occasion Daniel and Adam agree that it is fairer to judge on the performance itself. Daniel and Adam's exchange highlights the difficulty in deciding what is a fair decision. If Tanya was on the border between a pass and a fail would Tanya's effort have factored into the score decision?

The teachers invest time and thought into the AOP assessment and how it may, or may not, be working. Furthermore, this real-time reflection is an integral part of bettering assessment processes in the future. Kalthoff (2013) in an ethnographic study of grading practices in a German high school, found that 'teachers are involved in the process of assessment as generators of scholastic judgement' when they 'evaluate their own performance as well as that of the students' (p. 102).The teacher assessors capture what really matters while engaged in the process.

Our ability to enter into a dialogue and reflect on our own actions is a core part of being human and acting humanely. Hannah Arendt regarded thinking as crucial to ensuring that harm was not committed by individuals. She regarded dialogue as a crucial means by which we can safeguard our humanity:

> We humanize what is going on in the world and in ourselves only by speaking of it, and in the course of speaking of it we learn to be human. (Arendt, 1955, p.32)

This idea resonates with what I encountered in decision-making processes in AOP assessment. Through language our thinking becomes visible and open to challenge and progress. In rating

discussions, in critically reflecting on the decision-making process, teacher assessors shift the focus to the validity and fairness of interpretations and theoretical rationales regarding AOP assessment in their contexts. The social dimension of practices, including the consequences of assessment decisions, is brought to the fore in the decision-making talk (see wider discussion and evidence of this in 6.7 on The Social Filter). Human action and consequences are laid bare and in so doing the teacher assessors confront the nature and weight of their decisions. What the teacher assessors are communicating when they pause going through the motions is: 'we care about these decisions and we are endeavouring to take humane action'. The flickers of sustained agonized thought are to be cultivated so humane assessment decision making can thrive and be a prominent feature of assessment practice.

5.8   Chapter Summary

This chapter has presented the five core processes teacher assessors engage in when designing and implementing AOP tasks: Inheriting and Influencing AOP Assessment, Constructing AOP Task Representations, Communicating AOP Task Representations, Gathering and Selecting Assessment Evidence and Critically Reflecting on AOP Assessment Practice.

The analysis above demonstrates that the teachers influence assessment practice but to a lesser degree than teaching practice. Teacher assessors' AOP task representations were shown to be under constant construction and expectations linked to these representations are communicated explicitly and through more implicit nudges. The teacher assessors give cues to students on how to best fulfil the task prior to and during the assessment events. Evidence with which scores are reached comes from performance in the AOP task and from information garnered from outside the AOP performance. This includes performance on other assessments on the module and programme and from assessment opportunities in class, adding to the complexity of the EAP construct (this point is elaborated on in Chapter 6 and Sections 8.2 and 8.4). The teacher assessors engage in a process of discerning which evidence to admit and in some points in their practice engage in a critique of their approach to decision making. Teacher assessors episodically dedicate time and thinking power to evaluating their own practice. In doing this, they question, and aim to strengthen, the validity of the assessment (see discussion on validity and fairness in 8.6). The wider implications of such findings are taken up in Chapter 8.

With this theory as to the processes at work (before, during and after AOP assessment events), the next chapter takes a closer look at what is happening when teacher assessors are engaged in reaching score decisions and rating AOPs. The understanding of processes engaged in before, during and after the assessment events provides crucial insights into what influences the

implementation of a workable and valid or invalid assessment process, addressing RQ3. Most crucially, for the purposes of this study, an exploration of these processes aids analysis of what happens in real-time score-reaching decision making. The next chapter offers descriptions and theories as to how teacher assessors at the two sites reach score decisions and what they look for and value.

Chapter 6   Teacher Decision Making in Rating AOPs

6.1   Introduction

The purpose of this chapter is to present what teacher assessors do in the rating conferences. The analysis reveals what strategies and tools teacher assessors use when reaching and justifying score decisions which links to RQ3: 'What factors determine the EAP AOP assessment construct?' The analysis also uncovers what EAP practitioners look for and value in AOP assessment performances and what this indicates about the EAP assessment construct, therefore, addressing RQ2: 'What is the EAP AOP assessment construct?'

While time in the field has yielded data about what teachers look for and value from field notes, documents and interviews, in this chapter I primarily draw on data from the rating discussions. I have focused on this data set because it provides telling glimpses into the justifications behind assessment decisions at the time these were made. Interview data is also extensively drawn upon as it offers insights into possible motivations behind decision making which are not visible in the rating discussions. In Section 6.2, I give an ethnographic description of the AOP assessment event. The AOP assessment event refers to when students deliver their summative presentations to an audience of peers, two teacher assessors and myself. The remaining sections lay out the complex score-reaching decision making process which are represented visually in Figure 12 below. The Figure shows a picture of a student giving an academic oral presentation on the topic of Vygotsky's Zone of Proximal Development theory. The 'Zoom' label and close-up of the slide from the oral presentation represents how teacher assessors may switch from a holistic look at a presentation to a focusing on specific parts. Below the Zoom section are the 'Lenses'. Teacher assessors may judge, mark, describe, evaluate, or focus on technical or aesthetic qualities of, a performance. The visual then depicts the 'Filters' the performance may go through. The use of criteria, a ranking strategy and the social and (inter)personal influences affect how the AOP performance is assessed. The filters may overlay the lenses at various levels of intensity. For instance, a rater may judge a performance by ranking students and be influenced by social and personal histories. The final part of the graphic represents the end product of the process being the score which signifies a pass or fail mark.

The AOP

Zoom

Lenses

Judging

Marking

Describing

Evaluating

Technical

Aesthetic

Filters

Criteria

Ranking

Social

(Inter)personal

The score

| Fail | 45 | 46 | |
| Pass | 52 | 56 56 | 70 |
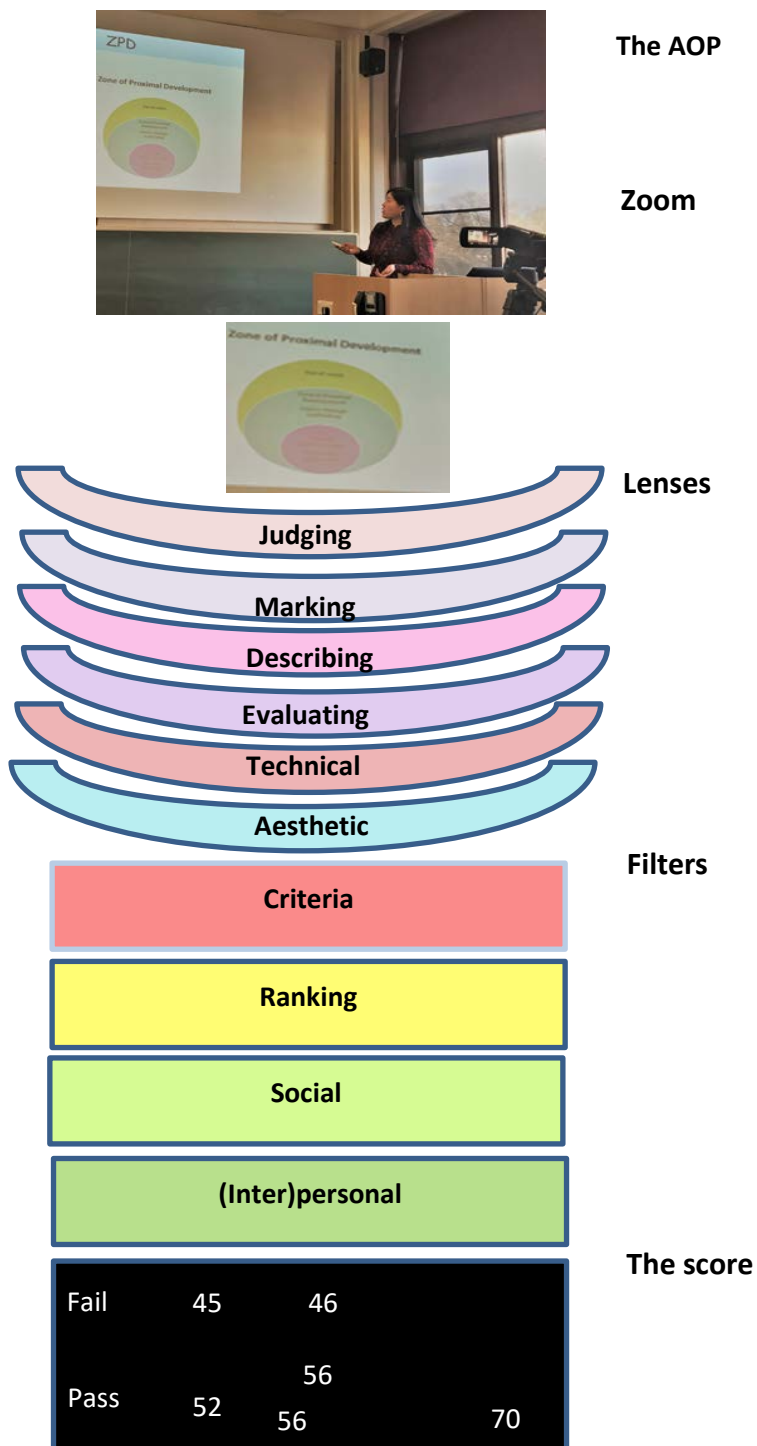
Figure 12: Visual Representation of AOP Teacher Assessor Decision Making

Section 6.3 presents analysis which evidences that teacher assessors use a zoom function to take holistic and close-up views of the performance. This is followed by Section 6.4 on the lenses through which teacher assessors process the performance. Sections 6.5-6.8 are then devoted to discussing the range of filters teacher assessors put the performance through. The processes are not fixed and linear as the representation in Figure 12 may suggest: the dynamic and varied use of the zoom, lenses, filters and multiple sources of evidence beyond the AOP performance will be illustrated in the preceding sections. Importantly, in the chapter summary (Section 6.9) a visual (Figure 17) which depicts the dynamism of the process is provided.

6.2   The AOP Rating Processes

At the two sites, prior to the AOP assessments, the teacher assessors provided task information through documentation and communications in class (as discussed in Section 5.5). In both of the sites, the students were given crucial information related to the task format, which I have summarised in Table 14 in Section 5.2.1. Rubrics were used during the rating process (see summaries of criteria in Tables 16 and 17 in Section 5.2.2.) I categorise the AOP assessment event and the rating discussions as the official assessment events and official sources of evidence for rating decisions. I describe the AOP events in the two sites in the next two sections.

6.2.1   The AOP Event

In FS1, the group presentations took place in one sitting. As the students arrived Daniel asked the students to sit with their groups at the front of the room. The teacher assessors planted themselves on the row behind the students set to the right. I positioned myself on the same row as the assessors a few seats away to the left of them. The students were excited and there was some nervous giggling at the sight of the camera. Adam stood up to announce the order of presenters, adjustments being made as one student would be arriving late, and Adam and Daniel went through a few 'house rules'. Adam and Daniel sorted their papers and had the rubric on the table for reference. During the presentations Adam and Daniel were making notes. The changeovers were swift with few words exchanged between the assessors during and in-between presentations. The final presentation was delivered and Adam closed proceedings in haste and rushed to another class. The presentations were scheduled on a Friday afternoon and Adam and Daniel planned to have their scoring conference to agree marks on the following Monday morning.

In FS2, there were two sittings: one in the morning and one after lunch. I was helping arrange tables with Georgina and Tracey when the first students arrived for the first sitting. The students were reciting their scripts and gazing at other students. There was a sense of nervous anticipation – a different atmosphere to the lessons I had observed. A few students started singing and making rather eccentric squeaking noises which they confirmed with me later were playful signals of their anxieties ahead of the assessment. The teachers tested the camera, Tracey arranged her papers with a military precision, and Georgina casually set up her laptop trying to find a convenient charging point. By this time the students were restless and impatient to start. Georgina passed around some sweets and the students looked visibly more relaxed. She gave supportive comments: 'we are not here to mark your mistakes, we are here to mark your content' (FS2 field notes). Georgina went through the order of presenters and the presentations began. Georgina and Tracey sat apart when rating. Georgina and I were on the same rectangular block of tables as the students and Tracey was on a separate table near to the camera to the right of us. Some students were asked questions and there was very little interaction between Tracey and Georgina during the presentations. The procedure was similar in the second sitting, however the atmosphere was more subdued.

6.2.2   The Scoring Conferences

In FS1, Adam and Daniel met to allot scores three days after the presentations. They referred to parts of the video recordings, their real-time notes and their memories of the AOP events, as well as events external to the AOP performances (as discussed in Section 5.6). Adam and Daniel discussed each student's mark for each of the five sections in the rubric consecutively but did not mention each criterion included in the descriptors for each student.

In FS2, the scores were allotted on the day of the presentations after having watched five presenters in each of the two sittings. Tracey and Georgina discussed each student often giving long individual turns about a mixture of aspects, traceable or not, in the rubric descriptors, based on their notes. They generally provided their interpretations of the performance and then put forward a tentative overall mark for the performance.

In both rating discussions, the teacher assessors take close-up and wider pictures of the performance, look at the performance through different lenses, and put the performance through numerous filters. In the next sections (6.3-6.8), I give details of the tools the teachers use exploring the dynamics, properties and the relationships between them. I begin with the zoom function.

6.3 The Zoom Function

In parts of the rating process, the teacher assessors zoomed in to take a closer look at an aspect of the performance in a moment of time or zoomed out to get a holistic take on the performance as a whole, or a holistic view of a specific feature of the AOP delivery.

Figure 13: The Zoom Function

Daniel highlighted the unreliable snapshot nature of some of the comments noted down during real-time marking of AOPs. He raised the concern that scratch notes on specific parts of the AOP performance could be misleading. When discussing a student's pronunciation, Adam and Daniel discussed mispronunciation of a word that they had both written on their notes.

> Adam: I'd written down 'questioneer' so I think a pronunciation of questionnaire was slightly
>
> Daniel: But didn't you think that was because it was generally her fluency and accuracy of pronunciation is quite good and then they'll be ONE WORD that is completely wrong and you kind of pick [on it]
>
> Adam: [Yeah] it's sort of variable isn't it
>
> Daniel: 'Cause I wrote down the same.
>
> {…}
>
> Adam: Well I've gone 10 or 11. But maybe I don't know 11 [top of that band.]
>
> Daniel: [I think we're] distracted by the 'questioneer'
>
> Adam: Ok 11 then. We shouldn't be distracted by one 'questioneer'!
>
> (FS1, Rating Discussion)

Daniel perceptively notes this example is not he feels an accurate reflection of the student's general fluency and pronunciation in the entire AOP performance. Daniel labels this proposed piece of evidence as a distraction. When making notes on close-up examples, the raters may be making note of the features which catch attention and not of those which reflect the overall standard of performance. In this case, the language item selected had the danger of giving the opposite impression of the presenter's performance regarding pronunciation. The assessors used a strategy of zooming in and out to check how valid their interpretations were of the performance.

The teacher assessors also move from holistic interpretation to close-up episodes as a way to consider how the overall impression of the performance may or may not fit with the specific examples taken from the presentation.

Adam: Accuracy I think is good but when you say when you look at it in close-up it's not as good as it seems

Daniel: But the key thing is that it doesn't affect understanding. You do this sort of translation you're constantly mentally correcting her aren't you. So I would say 12 for that would you?

(FS1, Rating Discussion)

Adam and Daniel consider the 'accuracy' of language used in the performance at a global level and in close-up. When taking a closer look, Adam believes his holistic take is flawed. Daniel considers whether this hinders intelligibility and decides that it slows processing time. This opens up contentious issues as what should be deemed inaccurate and how accommodating teacher assessors should be and to what extent they should focus on form (as discussed in Sections 2.9.2 and 8.5). There is an argument to be explored as to whether zooming in matters if inaccuracies are not detected at a broader level.

The teacher assessors are making a complex switch in cognitive processing while marking in real time and in their rating decision talk. This raises salient validity considerations (discussed in Sections 8.6 and 9.2.1). This is not the only tool teacher assessors summon and manipulate when watching and rating AOPs. As well as changing the level of magnification the teacher assessors employ different lenses through which to view the performance.

6.4   The Lenses

In the rating discussions in both sites, the teacher assessors view the performance through a range of lenses. They give qualitative judgements on the performance (judging lens) and then move to transforming those into quantitative measurements by scoring the performances (marking lens). In their rater talk, teacher assessors give descriptions of their experiences of the AOP performance (describing lens) and, in other instances, shift to giving a critical appraisal of the performance (evaluating lens). Parts of the decision making focus on the technical skills that the students display (technical lens). Other episodes include interpreting the performance's worth based on how pleasing the qualities of the performance are, according to the teacher assessors' sensibilities (aesthetic lens). The lenses may be applied to the AOP performance but also to other performances external to the AOP (as outlined in Section 5.6). The sources of evidence depend on what filters are used, specifically how strongly the criteria and social filter feature (see Sections 6.5 and 6.7 below).

The teacher assessors use and switch from these diverse lenses at various points when discussing the AOP assessments: there is not a fixed linear progression from judging to marking for instance. Below is an extract from FS1 rating discussion which is part of the teacher assessors' discussion of
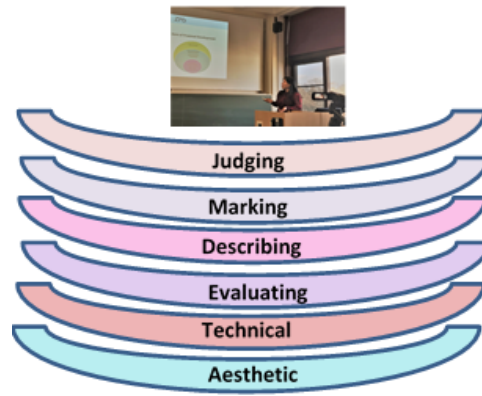
Figure 14: The Lenses

Anna's performance in her group presentation. I have divided the continuous utterance, admittedly rather crudely, into two parts – in part one the assessors mainly make qualitative judgements on the performance. In part two, the discussion shifts focus to marking (assigning a score). In the following data extracts, I demonstrate judging and marking lenses at work alongside the other lenses (describing, evaluating, technical, and aesthetic). The extract, which involves Adam and Daniel discussing Anna's performance on content and language, involves a complex integration of all facets of the decision making model (zoom, lenses, filters), but in the preceding analysis I focus on the lenses:

Part 1 – Judging

Adam: Anna I thought she was good, actually it's a thing for all of them no notes! Hardly anybody used notes! They were very confident I thought.
Daniel: They were very well prepared all of them weren't they.
Daniel: I thought she was kind of weak on content
Adam: Yes.
Daniel: But that's Anna.
Adam: Well yeah it's difficult with the content because it's kind of the whole presentation in a way. But I do think it did seem a bit weak. I thought also. I did write a note about language that she said things like 'people are very diversely' or something like that, 'because you know', she uses 'you know' a lot.
Daniel: It's interesting isn't it. I put that down. So she often makes mistakes with word form, but you don't focus on it because her delivery is incredibly confident. She's speaks at a good volume.
Adam: Yes, she's very clear
Daniel: Um and she's very confident. And then listen to the content and one she's making grammatical errors and two there isn't much there.
Adam: Yes, I think that's probably a good sort of summary. (FS1, Rating Discussion)

Daniel and Adam talk about the aesthetics of the performance (aesthetic lens) with the 'confident' delivery and technical details (technical lens) such as the level of volume and 'mistakes with word form'. The raters narrate what Anna does: 'she uses 'you know' a lot' (describing lens) and, in parts, signal whether particular features of performance are positively or negatively received: 'it did seem a bit weak' (evaluating lens). The teacher assessors describe content as 'weak' but do not describe what technical or aesthetic characteristics lead to this evaluation being made at this point in the discussion. The consensus between Daniel and Adam here is that when close attention is devoted to the content, as with the language, the cracks begin to surface despite a confident delivery. Whether the 'mistakes with word form' should affect the mark could be disputed in light of ELF research, however this is not the focus of this section. The teacher assessors continue their discussion of Anna's performance by considering what score should be assigned for the 'content' section in the rubric. They now view the performance through the marking lens:

Part 2 – Marking
Daniel: So I think I'd put her in the 9 or 10 for content.
Adam: Oh OK? I've gone 11, 12. But we could compromise. Sort of 11.
Daniel: Umm I don't think she's well-researched and there wasn't, I didn't feel there was much analysis.
Adam: Ok.
 Daniel: And the questions were handled confidently I'd say but I'm not sure if she actually answered [them.]
Adam: [No] It's interesting to see who sort of takes charge of the questions in the group isn't it. And she's one that would take charge hehe of questions! I don't know, would you go 10 or 11.
Daniel: 11. (FS1, Rating Discussion)

The teachers, when switching to assigning scores (marking lens), take on a whole new interaction ritual. Negotiation and compromise feature prominently in the dialogic communication (see Section 6.8 on The Interpersonal Filter for a consideration of the elements at play in dialogic

decision-making talk). In a number of instances, the descriptions become intensified and, in this case, reference to the descriptors in the rating scale are used. 'Well-researched', 'analysis' and 'questions handled well' feature in the rubric prose descriptors. The rubric is referred to in order to make a tentative score credible (see more on use of the rating scale below in Section 6.5, and specifically on its use in maximising accountability in Section 6.5.3).

The teacher assessors are constantly reviewing the performance by using an assembly of lenses. This extract also supports findings presented earlier (in Section 5.6) that teacher assessors in the two sites opt to not simply match productions in the performance with features listed in the criteria; the performance is experienced and assessed aesthetically in some cases with reference to aspects not contained in the criteria (e.g. confidence) with the response later aligned to criteria.

6.5   The Criteria Filter

In this section, I describe how assessors may, or may not, draw on the rubric in order to determine what to look for in the AOP performances when reaching score decisions.

In most sections in 6.5, I do not refer to large stretches of data as in other analysis sections. This is because considerable parts of prose descriptors needed to be redacted to best ensure a high level of unlinked anonymity, making the data extracts redundant. I, instead, describe instances of when there was stronger alignment with the criteria and episodes when teacher assessors use the rubric to aid decision making to a lesser degree. The teacher assessors refer to the rating scales, at varying levels, for reasons of alignment with task features and purpose, manageability, fairness and accountability.

Figure 15: The Filters

6.5.1   Alignment with Task

During my time in the field, Adam, Daniel, Georgina and Tracey shared their thoughts on the rubric's flaws and strengths. In the rating discussions, the teacher assessors refer to the criteria deemed relevant to the task, the specific performance, the individual and the social setting. By the teacher assessors' own admissions, work needed to be done to ensure the rubric reflects the goal of the assessment and task features.

Teacher assessors moved closer to or away from the use of the marking criteria depending on whether it reflected what they regarded as the assessment purpose(s). The prime purpose of the assessment was identified as ascertaining the students' readiness to embark on academic study at a UK university. Despite using the criteria to structure the discussion and heavily referring to it in parts of the decision-making talk, in FS1 and FS2 there were many instances when the talk shifted away from the prose descriptors. In FS2, the competences that Tracey and Georgina look for in the presentation can be traced in the rubric descriptors. For instance, Tracey and Georgina discuss features related to 'grammatical accuracy', 'pronunciation' and 'content'. However, the bulk of their detailed analysis is largely related to content, even though the criteria contain more sections devoted to linguistic features. In FS2, Georgina shared that most of the students' level of English was acceptable and would develop over time, but their academic potential which has been developed in classes, reflected in the content criteria, would need to be demonstrated to truly ascertain readiness.

The alignment with the AOP task, in particular the affordances of the AOP genre, strongly influenced what was discussed in the rating discussions. Selection and packaging of information suitable for an oral presentation format (e.g. use of text and visual graphics on slides) is mentioned numerous times in FS2 rating discussions but this is not covered in the rating scale. The teacher assessors also comment on aspects such as 'confidence', 'energy', and 'delivery style' which are also not included in the descriptors. In terms of aspects which could be categorised as presentation skills, Tracey and Georgina discuss these in greater detail than linguistic features. Georgina and Tracey give holistic, impressionistic judgements on language such as 'language-wise she's very good' without expanding on this. Similarly, in FS1, 'complex language' is a feature listed in the criteria but it is not discussed in the rating discussion. This strongly indicates that features of communication closely related to the academic oral presentation task override the criteria in the rubric if the rubric does not include such features needed for task completion.

The lack of discussion regarding 'complex language' may be explained by Adam's discomfort (expressed in interviews) with what complex language looks like in an AOP and his general discomfort with assessing accurate language use:

> Adam: With the accuracy and the range as we talked about before, I find it very hard to assess you know based on particular errors, and it really I think it has to be errors which affect intelligibility or it has to be kind of also just the range of language being used. I think it's quite hard to listen for complex language or complex grammar structures. (FS2, IAP with Adam)

Adam clearly states that it's difficult to listen for complex language in a presentation and in a later discussion with Adam he poses the question of whether many people use complex structures in

spoken utterances. He questions the applicability of such criteria to the task and also indicates this is difficult to notice. The difficulty noticing language was a key code in S1 and S2 data in general when watching real-time AOPs (validity ramifications are discussed in Sections 8.5, 8.6, 9.2 and 9.4). This points to issues with the applicability of the rating scale criteria and also the manageability of the task.

6.5.2   Manageability

The criteria is used as a means by which to increase the manageability of the task but at the same time seems to add to the teacher assessors' cognitive load.

In FS1, the criteria clearly provides a degree of structure and focus. For the most part, Adam and Daniel structure the rating discussion using the criteria sections: they discuss each individual's mark for 'content', 'fluency and intelligibility', 'accuracy and range' 'presentation skills' and 'group work'. In the extract below they say the key words of the names of sections in the rubrics to signal the structure. In the following exchange, the talk to follow should be related to 'accuracy' and then 'presentation skills:

> Adam: Accuracy. I think is good but when you say when you look at it in close-up it's not as good as it seems, it sort of seems
>
> {…}
>
> {discussion of accuracy descriptors removed}
>
> Adam: Presentation skills. Though I'd go a bit higher to be honest. I think she's confident and it's probably her strongest area.
>
> Daniel: Yeah. (FS1, Rating Discussion)

The rubric sections provide a tool for maximising consistency through the methodical application of these core categories to each student. In FS1, Georgina and Tracy use the two themes of content and language to structure their discussion which involves merging the linguistic section into a broader category.

The rating scales in both sites are also shown to be far too complex to be applied consistently in marking AOPs. The assessors are unable to discuss all criteria in detail due to the amount, and nature of some criteria with the time and resources available. For instance, in FS2, Georgina and Tracey merge descriptors into new, broader categories. In regard to content, Georgina and Tracey create a descriptive/analytical dichotomy which encompasses and ignores the range of the descriptors in the 'content' section.

It is in the high-stake decisions for students on the pass/fail boundary that the lengthier focused discussion of specific criteria takes place for reasons of accountability and fairness.

6.5.3   Accountability and Fairness

For reasons of accountability the teachers heavily referred to the criteria when assigning high or low marks. Adam and Daniel and Georgina and Tracey tended to move closely towards the criteria in critical episodes, such as during a high-stakes and controversial negotiation that may result in a student failing.

A critical episode in which greater alignment with the criteria was tangible is when Tracey and Georgina discussed a student that had been identified as weak. Prior to the AOP assessments, Georgina and Tracey pointed out a student to me in the café before the assessment and both told me how he was probably going to fail his presentation after lunch. I was a little taken aback, but having seen the close contact they had had with the students, I was not surprised they had an inkling as to who may not perform so well.

We watched the assessed presentations after lunch and Georgina and Tracey discussed each student's performances. The student they had pointed out in the café was next. Georgina opens the discussion by giving a detailed description of the presentation. She focuses on the issues with content and pronunciation and how the student has not acted on feedback. Georgina is forthright in giving her interpretations of the AOP performance and confidently questions Tracey's interpretations. Georgina is direct in her way of communicating in many settings and she is the main teacher of the class who has seen the students deliver practice versions of this presentation. Tracey gives her interpretation of what has happened and laments the fact the student did not use what he had learnt with her in classes. In the excerpt below, Georgina and Tracey continue this score-reaching dialogue by turning to the mark the student should reach (adopting the marking lens). Both the teachers may sound unfeeling and cut-throat in their talk in the extract to follow but this comes out of a place of deep disappointment and frustration as they were willing the student to have acted on feedback and to perform well. The extract begins with Georgina stating that the student for her is a fail but that she has not decided on a score. Tracey then proposes a mark of 50 on the border of a pass and fail. Tracey puts forward another tentative mark 5 points lower that is a definite fail. The teachers turn to the criteria and discuss prose descriptors for 7 turns to ensure that they are in the correct band. The discussion of criteria has been redacted as publishing this means there is an increased risk of the institution and informants being identified.

Georgina: So this is a fail. Clear fail, I don't know what mark.

Tracey: (sighs) I gave it I was between (sighs) I stuck a 50 there but I was kind of in between 48 and 50.

Georgina: I can't see a 50. Especially because at least one third of it I wasn't able to understand the way he was pronouncing the words and I was just saying to Louise {the researcher} some of the sentences didn't have an end.

Tracey: No I know (dejected tone). It's just such a shame because I've got {in notes} he could have covered marketing, market segmentation, promotion branding. ALL these things he could have picked up from his business module and fed into this!

Georgina: But just like, Tracey, just like with his essay.

Tracey: I know.

Georgina: He was never able to find [select]

Tracey: [I know]

Georgina: Analyse synthesize NOthing

Tracey: I know

Georgina: All year!

Tracey: I said to him this should be really good. You should be able to find HUGE amounts of information on {two company names redacted}. I mean the Business Schools set this out all the time. What do you want to go for? 48?

Georgina: I've got yeah 48.

Tracey: Alright 48.

Georgina: Even that is bloody generous I can tell you.

Tracey: Well you can drop it if you want to.

Georgina: Yes. [I do]

Tracey: [45?]

Georgina: Yes.

Tracey: I think 45 is actually probably more realistic about it isn't it.

Georgina: Yeah because (looks at notes indicating there are a lot of negative comments)

Tracey: (sigh) I know I've got them too (looks at paper full of comments) pronunciation is difficult to understand.

{Teachers refer to and read parts of the marking criteria verbatim for 7 turns. They do this to locate applicable criteria for reaching a score decision.}

Georgina: That's it thank you {directed at the marking criteria}. 45

Tracey: Yeah I agree with you. {…}

Georgina: It's going to be unfAIR to say to him that you'll be Ok (FS2, Rating Discussion, Part 2)

As found in similar ethnographic studies of marking practices (Kalthoff, 2013), when building an overall impression of the performance, the teacher examiners consider the score that the student requires to pass. Georgina seems to have a fairly clear idea of what a fail looks like but cannot instantly locate the student's performance on the scale. Tracey explores the possibility of fashioning a pass mark, but quickly ascertains that this is not feasible after Georgina's case is put forth. In this segment of the conference, Tracey and Georgina make reference to aspects at first unrelated to the criteria. These include the student's uptake of feedback and opportunities provided during the course and performance on other tasks (see more on the social factors which shape decision making in Section 6.7). Prose descriptors have been shown to be 'vague' and 'open' (Wilds cited in Fulcher, 2012) and 'what is scaled is teacher and rater perceptions' (North, 2000). The teacher assessors, however when reaching the final score return to the rating scale. It is physically brought closer to them and they read out significant parts of the prose descriptors in the band just below the pass mark to ensure a fail mark has tangible support. They consider how the AOP performance may fit these descriptors. These actions elucidate the fact that Tracey and Georgina align closely with the rating scale as an accountability strategy. They have identified, as Fulcher (2012), that rating scales have a function of holding teachers accountable, so they make sure it can be used to justify their overall decision.

This section (6.5) has demonstrated ways in which teacher assessors adhere closely to the rating scale and exercise distance from it. An indication that the teacher assessors do not deem the rating scale a sufficient tool in facilitating the score-reaching process is that the teachers, at both sites, compare individual student's performances with other students, rather than purely mapping the AOP performance to the rating scale. This use of comparative judgement pervaded the rating conferences and was identified as a core category and classed as a filter.

6.6   The Ranking Filter

A strategy of constant comparison is used throughout rating discussions at the two sites showing that the teacher assessors use criterion referencing and norm referencing. When assigning scores

to an individual student, the teacher assessors refer to the performances and tentative scores they have allotted to other students.

> Adam: Ok. Mm-hm. Um well 12 for presentation skills? Or 13 the same as Henry.
>
> Daniel: She wasn't as good as Henry.
>
> Adam: 12 then [or 11?]
>
> Daniel: [She was] there with Anna or Rochelle wasn't she. [I think we]
>
> Adam: [12 then.] I think she was better than Rochelle. At least she was more confident
> and that's basically a mark of confidence.   (FS1, Rating Discussion)

In this extract, Adam and Daniel assign a presentation skills mark for a student by comparing their general sense of the performance with the other students' performances, apart from an explicit mention of one quality: confidence. The assessors do not identify the presentation skills that the student has or has not displayed. This extract is an example of a movement away from criterion-referenced rating practices and an implementation of norm referencing in its place. Louis L. Thurstone 1927's work on the law of comparative judgment and subsequent research in language assessment conducted by prominent researchers such as Pollitt and Kimbell has highlighted the need human assessors have for a point of comparison. This use of constant comparison is a strategy which seems to be employed to ensure greater consistency of judgements and shows that, in instances, teachers instinctively feel more adept at making *comparative* judgements rather than purely criterion-based ones. Qualitative judgements made on Henry, Anna and Rochelle earlier in the rating discussion are somewhat detailed and references to the rating scale descriptors are made. But whether these descriptions are sufficiently rich enough to make the norm referencing valid remains ambiguous. The validity of the norm referencing practice would need to be ascertained by a closer analysis of the content of each qualitative judgements made on the individual students used as points of comparison (a potential avenue for future research).

The teacher assessors show inclinations to move away from mapping performances to a rating scale. This is seen most prominently by their wish to account for the social dimension of the teaching, learning and assessment processes on their module and programme. The next section demonstrates how the consideration of the whole social unit factors into the AOP score-reaching decisions. Teacher assessors engage in a process of putting the performance through a social filter part of which involves admitting other sources of evidence, other than the AOP performance, into the rating process. These pieces of evidence are gleaned from social interactions in the wider social ecology of their EAP programme.

6.7 The Social Filter

'The social' namely encompasses the interaction between the people in the field sites and acknowledges the contextual factors in the wider course ecologies. In Sections 6.7.1-6.7.2, I demonstrate that teacher assessors consider mitigating factors and discuss the students as rounded individuals rather than focusing solely on their AOP performance. They do this in ways which has the potential to maximise fairness. In Section 6.3, I share what the teacher assessors, as experienced practitioners, value the most in reaching score decisions for the AOP assessment task. The qualities the EAP practitioners look for are set out in 6.7.3.1-6.7.3.4.

6.7.1 Mitigating Factors

In the rating discussions, the teacher assessors use the flexibility of the open dialogue to discuss the context and actors. The discussion of the rating scale criteria is teamed with an acknowledgment of specific interactions between teacher assessors and students before the AOP assessment event. During the rating discussions the teachers (on a few occasions in FS1 and in a considerable number of occasions in FS2) share what guidance and opportunities they have given students to make visible any mitigating factors. For instance, students' abilities which were demonstrated in previous assessments are brought into the conferences. When reaching a score decision for Tanya, a student that Adam stresses has been trying to save her group's presentation, Adam declares the advice he has given her and advocates on her behalf. He points out that she is capable of producing fluent speech but that nerves were a decisive factor. He also divulges that he has given guidance on the problematic area.

> Adam: Fluency. She unfortunately got very nervous and she did her thing which is where she repeats and goes back and reformulates constantly which is what she does when she's nervous. That's unfortunate because she was very good at controlling that in the discussion seminars if you remember.
>
> Daniel: Yeah.
>
> Adam: It's something that I said to her beforehand, just sort of keep calm and try not to do that 'cause it is a real thing for her to go back and constantly reformulate and she got very sort of excited. So I think she was very very nervous. (FS1, Rating Discussion)

Adam acknowledges that the reformulation is an area that has been discussed with the student, and therefore it would not be wholly unwarranted to allow this to shape the score. In this example, Adam also strategically references a previous assessment when Tanya minimised reformulations. This episode shows that teachers look for ways to bias for the best based on previous assessment opportunities and declare the guidance they have given a student in rating conferences. The teachers seem to do this in order to consider whether reasonable affordances

172

should be made for students. A pronounced feature of rater talk in which the teacher assessors consider mitigating factors is a focus on the person, rather than on one performance.

6.7.2   Assessing the Person

The teacher assessors -- Adam, Daniel, Georgina and Tracey – highlighted the need to create a fair assessment process which they describe at times as one that does not allow bias and elements external to the AOP performance to influence the assessment decisions. The teachers, however, in the rating discussions, interviews, and participant checks acknowledged that their previous contact with students undoubtedly feeds into their decision making, even if slightly, when rating the AOPs. The teachers spoke with conviction about the rounded pictures of the students which they had built up over the course of an academic year. They simultaneously expressed a degree of guilt about their inability to remain completely objective in the process. There is a tension here as to whether subjectivity or objectivity is viewed as maximising fairness.

Excerpt 1:

Georgina: Not having blind marking is a good thing because then you know a person and you just don't want to judge them on the 10-15 minutes at the end of the semester. You know what they're capable of. So I think I take the whole picture into consideration. I can't just erase it like in Men in Black! Hehe! It's quite difficult. I want to be fair to everybody. (FS2, IAP with Georgina)

Excerpt 2:

Tracey: It doesn't matter what people say. You have this marking criteria to help you avoid bias but you still know you've taught them for an academic year. You still know at the back of your mind. You're marking them on the view of what you're seeing and what you know. And I defy any teacher to tell you that they JUST mark to a criteria! You know I'm going to let these students loose are they going to be able to cope? (FS2, Rating Discussion, Part 2)

Georgina in the exchange above makes the case that subjectivity can be a force for good. In the film Men in Black intergalactic agents sometimes need to erase people's memories with their handheld device called neuralyzers. In her reference to Men in Black, Georgina divulges that erasing all past events is not possible or desirable in educational assessment when teachers have prolonged contact with students. Instead of previous histories contaminating the assessment process, Georgina and Tracey view previous contact as enhancing the validity of the decisions as to how students will perform on their prospective courses. Tracey believes teacher assessors mark on what they 'see' and what they 'know'. Tracey seems to suggest that what is seen in the AOP performance is shaped by 'knowledge' of the student garnered over time. I would contend that it is impossible to fully 'know' a person and whether they will cope or not in a target domain. However, Tracey and Georgina strongly support using the holistic picture they have constructed in

small group and one-to-one interactions over a year and diverse assessment opportunities to aid the judgements made on the assessed AOPs which ultimately decide readiness to embark on academic studies.

The notion that assessing the whole person is fairer than making decisions purely based on one AOP performance is a value that permeates the assessment practice. This could be viewed unfavourably as teachers have not communicated that such informal assessment opportunities are sought and valued. The assessors looked for what the students can and have improved and also predicted how they would function beyond the EAP programme. The practitioners, through their experience of, and socialization into, HE practices, have constructed an understanding of what they feel is required to function in the target domain. In an informal discussion which took place directly after the rating discussion, which was recorded, Georgina and Tracey state that the rating scale is not wholly sufficient to fulfil their role as gatekeepers to undergraduate study.

> Tracey: I'm not sure any marking criteria is perfect. You know you're going, I think you take the marking criteria that you think is FAIR. It gives a structure doesn't it. But you're still bringing in your years of what you've listened to before. (FS2, Rating discussion, Part 2)

The teacher assessors not only bring in their interactions with their current cohort of students, but also what they have learned from their experiences over the years. Sadler (1989) suggests that through feedback teachers reveal their 'guild knowledge'. This knowledge builds a notion of what constitutes good, or what Tunstall and Gipps (1996) refer to as 'a notion of excellence'. In the two sites, this 'guild knowledge' shapes the construct of what being a student ready to enter academia looks like. There are deep questions pertaining to the sufficiency and accuracy of the teacher assessors' knowledge of the target domain. From describing and analysing rater talk, I endeavour to unpack what this 'notion of excellence' seems to be for the EAP practitioners at the two sites. From analysis, I detected that the teacher assessors as gatekeepers were working with a notion of 'ability to cope' rather than a notion of excellence. The notion differed across individual teachers, but a core construct of 'ability to cope' and its qualities that the teachers consider to be important has surfaced from close analysis of the rating discussions and interviews at the two sites.

6.7.3   Qualities of 'Ability to Cope'

The assessors look for the tools that students possess in order to function on their target academic course. The signals which indicate the potential is there to 'survive' in these environments are detected. These terms of 'cope' and 'survive' may sound somewhat dramatic and have negative connotations associated with them compared to the term 'readiness' which is often used in EAP literature. The teacher assessors stressed that the pass/fail decision ultimately

rested on gauging if a student had the *essential* skills to embark on their studies. Talking in terms of the minimal level of skills required was crucial because the teachers would not wish to deprive students of an opportunity to enter Higher Education unless deemed absolutely necessary. Georgina, in interviews, stated that nobody is 'ready' for academia when they enter an undergraduate course, including home students. She believed looking for survival skills was the fair approach in gatekeeping decision making. The teachers in the field sites cover qualities beyond the scope of coping because students have strengths in a range of areas well beyond the base level of readiness. The qualities included in the notion of ability to cope in 6.7.3.1-6.7.3.4 below are those stressed as most important. The analysis below gives an indication of the core qualities teachers look for but does not necessarily reflect the pass/fail boundary as most students passed comfortably in the two field sites. Future studies which analyse what factors into decisions on the pass/fail boundaries would be required to establish clearer notions as to what constitutes the base level of readiness to embark on studies or ability to cope.

The teachers make reference to people, practices and scenarios on target courses informed by snippets of information they have gleaned over the years -- for example from working on in-sessional courses -- to justify their ideas on what the students require. In both sites, when wishing to strengthen an argument when there is some disagreement, or when justifying a low mark, the teachers made reference to beliefs about what the student may encounter and how they may function on their destination course.

> Adam: It's nerves
> Daniel: It's nerves and I think as a Master's student in a seminar she'll be very impressive after a [few months.]
> Adam: [Yeah.] Yeah OK let's put 12 then. (FS1, Rating Discussion)

The assessment is a presentation task but the performance in a seminar situation is taken into account. The teacher assessors suggest that from their assessment of Fiona on the course that she will shine in a seminar on her prospective course after having time to settle onto her prospective course. This demonstrates that projections as to how the students would function in the destination disciplines aid decisions as to whether the gates should be opened or remain closed.

The construct of 'ability to cope' may remain tacit in the data in many instances, however in some episodes the tools which make up this construct are articulated and clearly detectable. In rating discussions, feedback provided on practice presentations, guidance given in class and from interviews with the practitioners, there were features which were clearly stressed as important. I paid particular attention to when, how often and in what tone these features were mentioned. The next sections lay out some of the qualities of 'ability to cope' generated from a grounded analysis of rating discussions and interviews. These qualities are what teacher assessors value and

are not predominantly shaped by the rubric, therefore are included in The Social Filter. The codes which were generated from the data were divided into categories with some difficulty. Mercer et al.'s Oracy Skills Framework proved useful in refining the categories in the next sections (6.7.3.1-6.7.3.4) which had already been created from inductive coding but required greater clarity and precision. I adapted the categories in Mercer et al.'s framework and new categories were constructed.

6.7.3.1   Cognitive Skills

The word cognitive has been used for some codes generated under 'ability to cope' as these skills primarily seem to involve concerted mental action and had strong parallels with those within the cognitive section of Mercer et al.'s Oracy Skills Framework (see Figure 3 in Section 2.2). I recognise, however, that these skills are also shaped by socio-cultural factors.

The teacher assessors, in reaching score decisions, mention cognitive skills such as content selection and reasoning skills. The qualities of these skills include the students' ability to select, organise and assemble information; develop logical and sound arguments; and begin to take critical stances on issues. Showing signs of academic potential, or as Tracey puts it 'academic nous', are clearly very high on the teacher assessors' agendas. Cognitive skills which display this 'academic nous' include ones such as using reliable sources (in FS1), referencing appropriately, building logical arguments and repackaging of data in a digestible format (in FS2). Another core skill which is included under cognitive as it relates to 'task focus' is the ability to develop an appropriate mental interpretation of the task. This skill is shaped heavily by social-cultural factors, and has discernible overlap with the three other categories under 'ability to cope': communication skills and social and emotional skills and formative learning skills.

A student's understanding of what the task entails and what the assessors expect is a feature which was described as important by all assessors. In the data, impressionistic judgements of the performance were provided. Impressionistic evaluative comments included: 'it was good', 'a good presentation'. In certain parts of the rating discussion, each of the assessors at the two field sites comment on the student's ability to give a presentation in line with their conceptualisation of the task:

Excerpt 1:

Adam: Hmm. She is good at presentations I think. She was very good in her practice presentation. I think she comes out of herself.   (FS1, Rating Discussion)

Excerpt 2:

Daniel: He got what a presentation should be I thought.   (FS1, Rating Discussion)

Excerpt 3:

Tracey: I thought it was a really good presentation. I've put I really liked the presentation. It was a good style, calm, considered.   (FS2, Rating Discussion)

Excerpt 4:

Georgina: Um references are great and you know it was like a proper competent presentation.
                                    (FS2, Rating Discussion)

The teachers declared in the rating discussion that they appreciated a student's performance which aligned with the teacher's idea of a 'good presentation'. For each of the assessors, however, these ideas may be similar or different (as discussed in Section 5.4).

In FS1, the importance of shared task representations between student test takers and teacher assessors (specifically ideas on what an academic oral presentation looks and sounds like) was illuminated when one student did not meet the teachers' expectations and had for them an 'odd' presentation style.

Adam: Hmm, but I do wonder whether it's 'cause it is heightened when he's doing a presentation. I wonder how much it {the delivery style} is linked to his presentation style. Or what he thinks should be a presentation style. (FS1, Rating Discussion)

Ben, a student informant in S2, divulged in interviews that he had chosen not to prepare a script or rehearse his presentation because he wished to develop his ability to speak English spontaneously. He thought speaking naturally with only a loose familiarity with his slides acting as prompts would provide the best formative learning opportunity. Ben's understanding of the task is discussed more fully in 7.2.2 on student agency). The performance was incoherent and uncontrolled from the assessors' and my own perspectives, and Adam astutely recognizes it is linked to Ben's ideas on the presentation style. Through interviews with Ben it seems that he

holds different ideas on the acceptable level of spontaneity in a presentation to the teacher assessors. The student's ability to understand the task requirements and perceived core qualities of the genre were a key part of doing well. This echoes Wolfersberger's (2013) findings and supports his calls for task representation to be part of classroom-based assessment constructs. The communication skills - the second quality of ability to cope - displayed were a crucial part of the poor reception of Ben's presentation by teacher assessors.

6.7.3.2   Communication Skills

In Mercer et al.'s Oracy Skills Framework linguistic skills such as 'vocabulary' are separated out from physical skills such as 'gesture and posture'. From the analysis of data on AOPs, teacher assessors are shown to value the integrated use of physical skills and linguistic skills with other skills involved in AOP communication (such as 'pace of spoken language in conjunction with slides'). According to an analysis of the data which centred on gauging what teacher assessors truly valued, the 'communication skills' category involves use of spoken and written language alongside other modes of communication, organised in such a way as to convey the message with clarity and encourage audience engagement. I define communication skills as linguistic and non-linguistic modes of communication involved in delivery – it includes range of language and graphics in visual aids, structure and clarity of talk and of visual aids, but does not include aesthetic features such as visual design of slides and levels of confidence, flair and effort; these are covered in the social and emotional category (6.7.3.2). The skills involved in communication skills include: range of vocabulary and grammatical structures, logical and accessible structure of talk and visual aids, appropriate pace of talk in conjunction with visual aids, use of gesture, posture, facial expression and eye contact that supports rather than detracts from the key message, accurate use of language on the slides, and use of all modes of communication to best relay the message.

The teacher assessors included comments in the rating discussion and in lessons pertaining to the communication of meaning and audience engagement. It was not sufficient that the message was clear, the teachers required a degree of effort to be made to retain the attention of the audience. The assessors made judgements on the functional communication of meaning as well as the aesthetic quality of the communication. In Excerpt 1 below, Adam's comment on one student's presentation in FS1 encapsulates the emphasis assessors placed on feeling at ease during the

presentation and accessing the content, suggesting a relationship between the two. In excerpt 2, Daniel expresses how he looks for language use that does not jar or distract:

Excerpt 1:

Adam: I find it very difficult to follow and I found it sort of uncomfortable listening to it.

(FS1, Rating Discussion)

Excerpt 2:

Daniel: You're constantly making this adjustment to correct it as you listen to her and I find that distracting.   (FS1, Rating Discussion)

A key and thorny issue here arises as to what onus lies on the audience in terms of facilitating successful communication. Audience members, as well as the presenters, have a responsibility to maximise the potential for mutual understanding. In academia, I have witnessed established academics putting a heavy onus on the audience to digest material, and sometimes this is part of the education. In an EAP assessment context the judgements made on ease of communication is complicated by the fact that the assessors need to ascertain whether the student test takers are likely to be able to communicate successfully with those who do not have the opportunity to have prolonged exposure to their communication style and may not be accommodating interlocutors. Teacher assessors seem to try to gauge the strategic competence of students to cope with a range of situations and prioritise intelligibility. However, the social and emotional dimension may be crucial in heightening engagement, which is discussed below.

6.7.3.3   Social and Emotional skills

The social emotional skills valued by teacher assessors in determining students' 'ability to cope' and 'readiness' to embark on their studies involve aesthetic qualities of speech and physical movement, and visual aid design. Character traits (e.g. resilience, compassion), and skills involved in working with others (e.g. not dominating a Q&A discussion or unproductively interrupting the questioner) are also valued.

In numerous instances in rating conferences, the teachers commented on how pleasing the performances were aesthetically. The teacher assessors in FS1 describe the students' style of delivery using adjectives such as 'confident', 'engaging', 'perky', 'overdone', 'distracting', 'odd', 'serious', 'dull', and 'repetitive'. In FS2, presentations which were delivered in a 'calm', 'considered', 'moderated' and 'interesting' way were valued by Tracey. In both sites, the assessors became animated in the discussion when talking about the visual artistry and creativity that AOPs afford:

> *Excerpt 1:*
>
> Tracey: TELL you one thing I did like though, I liked her slides. Very clear. I liked the colour coordination. You know me [I like]
>
> Georgina: [hmm]
>
> Tracey: PowerPoint presentations. (FS2, Rating Discussion, Part 1)
>
> *Excerpt 2:*
>
> Daniel: One thing I would say about this group is that I thought the design of the slides was very good. (impressed tone)
>
> Adam: And I know that's down to Linda because she's a designer and those are Kristina's photographs. And the other thing I thought and have written down here is the design matches the content! They matched the idea of the circle. (FS1, Rating Discussion)

Commentary on the sensory experience of the presentation was not only confined to the graphics. The assessors comment on the energy that exuded from the student test takers. 'Conviction', 'confidence', 'cooperation' and 'resilience' were all mentioned as desirable qualities the assessors wished the presenters and their creations to embody.

The presenter's conviction and presentation style was considered to be a crucial factor in relaying the message effectively. The assessors revealed that they valued a student who considered the way that information is presented and showed awareness of what would engage their audience.

> *Excerpt 1:*
>
> Tracey: It's similar to what we said this morning isn't it. I've put shame you read your notes, it affects your pronunciation, you lose your sparkle. (FS2 rating discussion)
>
> *Excerpt 2:*
>
> Adam: And she's not putting a lot of energy into it, is she?
>
> Daniel: You really have to concentrate hard when somebody is as dull as that, don't you.
>
> Adam: Hehe, hehe!
>
> Daniel: I'm not being mean. But that's what you want, they should make you want to listen!
>
> Adam: Yes, that's true. (FS1, Rating Discussion)

What is considered pleasing communication in oral presentation tasks and its relevance to the particular assessment context needs further clarification. The effect on the marks seemed to be minimal and difficult to gauge, however, one social and emotional quality which permeated the decision making was 'confidence'.

Confidence was shown to be a feature which can be positive or negative: part of the substance of a performance or a superficial quality which does not count for much. In FS1 and FS2 teacher assessors indicated the importance of building students' confidence as a key driver in their teaching behaviour. Confidence in the eyes and ears of the assessors took on a range of forms. One manifestation was 'confidence with content', which referred to having a strong understanding of the content. A second manifestation was 'confidence in the delivery style' which relates to presenting in an assured way through voice, body language, and presence. A third manifestation was 'a quiet confidence' which involves a more introverted student demonstrating what they are capable of doing in an unflashy way. Confidence was depicted as a necessary quality for students to possess in some form in order to showcase their skills in AOP (and other) performances on an educational course. Confidence surfaced as a fundamental notion in guild knowledge, but often it is not possible to translate how possession of confidence affects the scores. In the next example from FS1 however, the effect of confidence on the score was explicitly shown in the allocation of marks – a student received an extra mark for their confidence in delivery:

> Adam: 12 then. I think she was better than Rochelle at least she was more confident and that's basically a mark of confidence. (FS1, Rating Discussion)

In the extract above, the teacher assessors use their strategy of comparing students to consider the consistency of their marks (as discussed in Section 6.6 on The Ranking Filter). Adam and Daniel believe in this case that the student has earned an extra mark for a confident delivery. While confidence is valued, it does not eclipse the substance of the performance:

> Daniel: I think she's a student that sometimes needs a bit of very pointed academic advice don't you. That she can't get away with being confident. (FS1, Rating Discussion)

Qualities such as energy, enthusiasm, resilience and cooperation were appreciated by the teacher assessors. These seemed to have the potential to bump up the mark only very slightly. Assured and confident delivery is appreciated but not regarded as an indication of quality in the AOP assessments by the teacher assessors at the two field sites. Confidence to perform and showcase talents is shown to be a necessary component of the teacher's notion of what matters though, but this can be achieved in multiple ways – confidence with content and confidence to showcase individual strengths. The students' effort exerted into reaching their potential is an element that the teachers look for above confidence. The teachers look for how students learn and act on feedback; they gauge the students' formative learning skills.

6.7.3.4 Formative Learning Skills

In the previous sections on qualities of ability to cope (6.7.3.1-6.7.3.3), I referred to Mercer et al.'s Oracy Skills Framework to aid the development and communication of categories. The category 'formative learning skills' is not based on a category in the Oracy Skills Framework. Formative learning skills refer to students' ability to notice and seize opportunities to learn on their educational courses. The students' engagement with teachers' instructions and feedback and students' willingness to learn are qualities valued in the two field sites. The teacher assessors underscore the importance of these skills in the students' ability to cope on their academic courses.

Georgina believes that the feedback given to students in practice or formative tasks is something to which students should be receptive and that the role of the student is to make the effort to understand requirements through feedback.

> Georgina: So if a student takes feedback on board that should be included in the mark I think. I think that for me that's very important that people show that they've understood what the requirements are. (FS2, Rating Discussion, Part 2)

In both sites, the teacher assessors shared with each other and myself who they believed would perform well and not so well. They were eager to communicate who had not attended lessons and the effect this may have on their assessment outcomes. The teacher assessors engaged in conversation regarding how much effort the students made and what they had done with opportunities afforded to them throughout the course when reaching high-stakes decisions. In FS2, Georgina and Tracey drew on knowledge of what support and feedback a student had received in order to ensure a fail mark was justified and justifiable. The excerpt from this deliberation demonstrates that the perceived lack of awareness and effort on the student's part informed Tracey's score making decisions:

> Tracey: He didn't take ANY of the stuff he'd learnt in business and bring it in to this presentation. I mean he's just taken a whole module of Business with me. So it's been a very naïve. You know he's never going to survive. It's just such a shame because I've got (looks at notes) he could have covered marketing, market segmentation, promotion branding. ALL these things he could have picked up from his business module and fed into this!
>
> (FS2, Rating Discussion part 2)

The assessors both agreed that the student had received the guidance and opportunity to showcase the skills expected. Both openly stated that they did not believe that he would 'survive' on his prospective destination course. Georgina, as the main teacher of the group, who had given feedback on two practice presentations, outlines and scripts, places particular emphasis on uptake of feedback and improvement in the AOP assessments.

> Georgina: So Nadia had feedback on her practice presentation. Everything I told her, asked her to do she's done. Um, it was improved. (FS2, Rating Discussion, Part 1)

The teacher assessors in FS1 also noted when students had displayed a level of investment in their studies:

> Daniel: Kristina is a student that is getting stronger week on week, isn't she? In her reading project she prepared. She's trying really hard.

Adam, also made similar comments about the importance of taking feedback on board, but was keen to stress (in participant checks) that this should not become 'a teacher always knows best' approach. The students need to be afforded opportunities to offer alternative visions and different paths to excellence.

The rubric did not include effort, reaching potential, uptake on feedback and progress in learning in the descriptors in either of the field sites, however these are themes which were continually revisited at both sites. While it is seen as a vital part of learning and discussed in rating discussions, on some occasions, the assessors at both sites (in some cases reluctantly) decided not to allow effort to determine the mark.

> Georgina: And so I mean she did a lot of reading and um so that doesn't really mean to say oh she's working really hard and that's why we're going to give her this mark. I know that's wrong.

Georgina may have communicated that taking effort into consideration on this occasion was 'wrong'. However, Georgina and other assessors also stated that a collection of assessment opportunities and a student's level of effort in the learning environment were crucial insights required for making the fairest decision possible (as discussed in Section 6.7.2 above). Georgina in an informal discussion with Tracey and I stated that she intends to make improvement and uptake on feedback a core criterion in future assessment.

> Georgina: Yes, see that's something that is not in the marking criteria BUT I think feedback and just ignoring feedback uff what's the point in me giving you feedback and how are you going to learn. They've done their best. They've tried so there you go. That's another tick. Well done student!

(FS2, Rating Discussion, Part 2)

The tension, for Georgina, exists as to what is perceived as fair in assessment practice in the current assessment task and what is required in the target domain. However, the real ambivalence is whether effort and uptake on feedback is a core part of the valued construct assessors are working with of 'readiness' or 'ability to cope'. The students need to display a standard of performance: effort without results is not sufficient. However, on EAP courses when assessments may have low predictive validity due to the lack of subject specificity, effort that leads to a positive or good enough result can be a helpful indicator of the student's impetus to cope, of their persistence and perseverance. This quality may stand them in good stead no matter what they encounter. In this sense, it is a key quality that will contribute to success in the academic realm, assuming the impetus to make effort on their target course is similar to that displayed on the EAP module.

The qualities which constitute readiness to embark on academic study are numerous and at times nebulous. However, clear priorities emerged across the two sites and among the 4 teacher-assessors. They value cognitive skills, communication skills, social and emotional skills and formative learning skills.

The aims of the two modules are 'improving English for masters/university study' The learning outcomes include some aspects of 'ability to cope?'. In FS1, cognitive skills and communication skills generally, and a couple of specific socio-emotional skills (teamwork and confidence), are explicitly stated in the learning outcomes in module outlines. In FS2, cognitive skills and communication skills are also described broadly as target skills in module documentation. Socio-emotional skills are also included in the aims of the module, with one aim being to increase confidence in academic contexts. How the core skills are reflected in module outcomes and the assessment may need to be revisited to ensure greater constructive alignment.

The implications of what the teacher assessors look for and value for the construct of EAP is considered in Section 8.4 and issues of validity in Section 8.6. The make-up of this construct of 'readiness' or 'ability to cope' varies between practitioners and is constantly evolving through new experiences and interactions with fellow practitioners, but the key areas of convergence have been demonstrated.

Personal and interpersonal factors no-doubt filter the score-reaching decisions and we turn to this next.

6.8   The (Inter)personal Filter

The location of the interpersonal filter at the bottom of the visual representation of the rating process (Figure 12) presented in the introduction and in the final part of this chapter does not indicate that interpersonal positioning is a weak phenomenon in the score-reaching process. Quite the contrary, in line with Widdowson (2020, p. 90), I see interpersonal and personal positioning as discourse. The interpersonal filter is a constant overlay through which events are filtered both in terms of how they are experienced and communicated.

Identities are brought into the assessment proceedings. The personal background, interests and values of individual teacher assessors shape their priorities and interaction styles. The interaction between individual assessors in rating conferences creates a unique and dynamic forum which shapes the tone and focus of the discussions. In this section, I first discuss the personal dimension of practice, before then turning to the interpersonal elements.

6.8.1   Personal Dimension

The personal and professional experiences of individual teacher assessors influence their approaches to assessing AOPs and their students. Adam's time giving presentations on English literature developed his keen interest in the aesthetics of presenters' delivery styles. Adam was an IELTS examiner for over 10 years and this seems to give him a firm grasp of the linguistic descriptors in the rubric. He has actively engaged in dialogue with fellow EAP practitioners and academic literature on intelligibility, including work in ELF, and this informs his view that intelligibility is the core goal in communication. Daniel strongly values critical thinking and looks for the students' ability to question. He laments about the post-truth era he believes we are living in and asserts that education is crucial in developing the skills to question information. Daniel

shared that he focuses on effectiveness of presentation technique in delivering the message also.

Georgina has a personal interest and professional specialism in the role of technology in learning. She has little experience of delivering presentations and has a slight aversion to the overuse of PowerPoint slides as she feels they can distract from the message. She increasingly places less emphasis on presentation style throughout the course (as discussed in Section 5.4.2). Tracey is currently studying for a doctorate in Education focusing on second language learner motivation. Tracey has had experience teaching Business modules. She looks for a presenter who has a 'professional manner' as she believes this impacts on the perceived credibility of the speaker.

These pieces of information about the personal background of the teacher assessors, are but the beginnings of a vast tapestry of threads that form their identities which shape assessment practice. The interactions between the teachers add further complexity to the assessment rituals and all interaction we encounter as humans.

6.8.2 Interpersonal Dimension

When the teacher assessors engage in discourse related to the AOP assessment decisions, the complexity of the processes at work become heightened with another layer of dynamic influences at work. The teacher assessors engage in processes of positional convergence and divergence. One key aspect of group decision making is the negotiation, or lack of negation, involved. In the following extract from FS2, Georgina and Tracey, through humour and metaphor, acknowledge the bargaining nature of the scoring decisions.

Excerpt 1:
Georgina: So anyway um . . . can we go with 60? 59? (hesitant tone)
Tracey: [I mean I did] (high pitch tone indicating doubt)
Georgina: [I mean she was] Because I did give it a bit more actually 'cause I thought the second half got better. I gave it 62.
Georgina: Yeah I'm happy with that.
Tracey: Yeah? Ok. 62 then hehe!
Georgina: Ok!
Tracey: Hehe! Going going gone! (thumps hand on the table imitating gavel at auction)
                                        (FS2, Rating Discussion, Part 1)
Excerpt 2:
Tracey: [How about] 78, 78.
Georgina: Right.
Tracey: 78. I think she deserves [that]
Georgina: Sold!
                                        (FS2, Rating Discussion, Part 1)

The teacher assessors tentatively put forward marks on many occasions and negotiate from this initial anchoring. Georgina puts forward a mark of 60 or 59 with hesitation and upon hearing

Tracey's high pitch tone goes up two marks and provides a justification. Unlike most bidding in auctions which increase in increments, the figures can go up or down when teacher assessors consider marks. Tracey's use of the phrase 'going, going, gone!' and gesture of a gavel hitting the table are synonymous with auction proceedings and specifically the closure of a sale in biddings. This metaphor highlights the negotiation involved in score reaching and also highlights the frenzied atmosphere of a rating discussion. The teacher assessors are tasked with assigning multiple scores to individual students within a restricted timeframe, which is similar to a large amount of lots to be auctioned in any one sitting at an auction house.

During the rating discussions the assessors manage a triangle of interests. The relationship between teacher assessor and co-assessor and relationships between teacher assessors and students. The teachers were mindful of the impact their behaviours may have or have had on their current and future interactions with co-teachers and students.
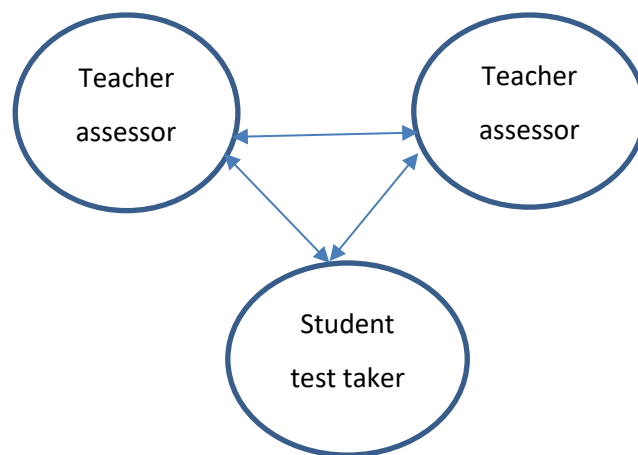
Figure 16: Triangle of Interests

6.8.2.1   Teacher-to-Teacher Interests

In interviews, Adam shared how he was aware of the diverse rating styles of different co-assessors he had worked with in his teaching practice and recognized the influence that this had. Adam was particularly attuned to the actions of his co-assessor and openly shared his willingness to be a cooperative rating partner in the rating discussions themselves. Adam also divulged how he was embarrassed during the discussion for fear of losing face, an honest admission revealing much about the interpersonal dimension of score-reaching dialogue:

> Adam: Yes I put 12 for that. Yep. No I was slightly embarrassed because you said she wasn't so good and I was oh maybe I've gone a bit high. But actually, no I agree. Yes, it definitely doesn't impede. She doesn't make mistakes as such. She doesn't make mistakes that impede communication.
>
> (FS1, Rating Discussion)

The rating decisions were shaped by the need to come to consensual decisions. While reaching score decisions, the assessors put forward tentative marks and evidence, often indicating that they were willing to adjust the mark. Showing the willingness to compromise on some decisions gave raters the opportunity to take a strong position in other cases. The dialogue reveals a tension between what Widdowson (2020, p.12) describes as 'the co-operative and territorial imperatives'. Widdowson describes all communication as involving a tension 'between the social

need for participants to relate to each other and the impulse to defend their individual space.' (p.12). In assessment exchanges, the teachers find ways to display and signal cooperation and to assert their identities as good teachers and accomplished decision makers.

The relationship between the assessors was not always harmonious, with imbalance in the discussion because of interruptions and dominance on behalf of one rater. In FS2, at times, there was a jostling for position during the rating discussion. The insider advantage of the main teacher of the group was palpable in the score-reaching talk between Georgina and Tracey. Georgina's time with the group, when she had marked scripts and seen two practice presentations, allowed Georgina to dominate parts of the discussion. Tracey however, drew on her time with the students in other modules on the programme to go some way in redressing the imbalance. Tellingly, rather than providing a more objective perspective, prior interactions with students gave teacher assessors a particular kudos in the rating discussions, in places.

The segment from the rating discussion below shows how Tracey endeavoured to give her account of a student's performance but Georgina continued to interrupt Tracey after having already spoken at length. Tracy resorts to employing robust rhetorical tactic to be heard using paralinguistic features of increased volume 'MINE!' and imperatives delivered in a firm tone 'Listen!', to stop Georgina stifling her ability to assert her views and ultimately carve out and protect her territory. Prior to this segment of data below, Tracey had made unsuccessful attempts to give her account, and Tracey laughs because of the length of time that Georgina has been talking. When Tracey finally was able to speak, Georgina interrupted and Tracey directly told Georgina to listen to her interpretation. 'Ok right' signals Tracey's third attempt to speak.

Tracey: Ok! [Right]

Georgina: [And] that was it!

Tracey: Ok MINE then. Ok! This could have been [a really]

Georgina: [COULD] have been! (high pitch)

Tracey: No listen (firm tone) this could have been a really good PowerPoint presentation. On his referencing on the slides on his reference area

Georgina: Hmm.

Tracey: He had some really good references there.

Georgina: Where?!

Tracey: Wait! What he didn't do. He didn't take any of the stuff he'd learnt in business and bring it in to this presentation.

<div align="right">(FS2, Rating Discussion, Part 2)</div>

Georgina and Tracey, in this exchange, were talking about a student who they had identified (prior to the assessment itself during lunch and at the start of the rating discussion) as somebody who was a weak student. This may explain why Georgina wished to give a full and detailed account and expressed disagreement so quickly with Tracey. Georgina seemed not to be concerned with conflict avoidance and protecting Tracey from losing face. Georgina and Tracey had worked together before and were comfortable challenging each other, but Georgina seemed more at ease with a direct communication style than Tracey. Georgina's knowledge of the previous presentations gave her an edge in the scoring discussions as she was able to point out aspects (e.g. inappropriate sources) more easily. This may explain why Georgina's interpretations were delivered with conviction, had more sway and Tracey was more amenable to compromise. I am acutely mindful that without knowledge of the context and previous happenings on the module, Georgina could be viewed unfavourably from reading the extract above. While Tracey perhaps should not have had to battle to be heard, Georgina was expressing frustration out of disappointment concerning how the student performed and controls future interruptions in the rating discussion. Both Georgina and Tracey verbalize their disappointment and concern for the student later in the discussion. The extract above demonstrates that rating discussions are suffused with emotion and can be highly tense encounters involving cooperation and conflict.

As well as managing teacher-to-teacher relationships, the teacher assessors consider their previous and future interactions with the students when reaching score decisions.

6.8.2.2   Teacher-to-Student Interests

Not only do teachers refer to interactions with students prior to the AOP when making score decisions (as discussed in Section 5.6.2), the practitioners also considered areas which required delicate communication when giving feedback to student test takers after the AOP.

> Daniel: Shall we put him in the middle of that descriptor then?
>
> Adam: No I think 12 is enough. I think it's more of an area it's awkward.
>
> Daniel: Yeah.
>
> Adam: Because he thinks that he's better than he is at that and I don't know that he is. Um because of his background. (FS1, Rating Discussion)

In this instance, the teacher assessors recognised that communicating the score on fluency and intelligibility to the student may be 'awkward'. This is because the practitioners believed the student had a different interpretation of his own level. This is explained by Henry's 'background', which is shorthand for Henry's experience as an undergraduate at a university in the USA. Because he has lived in the USA and studied through English medium the teacher assessors express concern that Henry may be particularly sensitive to comments about his level of fluency and intelligibility. This does not seem to shape the mark but the extract reveals that concerns for the teacher-student relationship are also racing through teacher assessors' minds and shape score-reaching dialogue.

The actions of teacher assessors were framed by a desire to have discussions which achieved a consensus and minimize conflict but also to exert enough influence so as to earn respect from their colleague. The teachers engaged in, or disengaged from, conflict and face saving strategies for the benefit of themselves, their co-assessors and students. A closer investigation of the interpersonal dimension would be an interesting and important avenue in future research.

6.9   Chapter Summary

The assessors use a complex and sophisticated mix of strategies to manage the assessment process. They use a zoom function, look at the performance through a range of lenses (marking, judging, describing, evaluating, technical, aesthetic), and put the performance through a range of filters (criteria, ranking, social and (inter)personal) at any given time. These resources, created and applied, shape how decisions are made and ultimately the assessment construct, and are key to answering RQ3. Figure 12 in Section 6.1 depicts these resources which can be drawn upon in reaching a score. In reality, these tools, strategies and influences (zoom, lenses, filters and sources of evidence) would not be used simultaneously at any given few seconds in the rating process. Different resources would be used, removed and then applied to the task again, whizzing in and

out of view. Some filters would be applied to a small extent, others discarded with multiple sources of evidence, other than the AOP, floating into the picture. This dynamic process is perhaps better depicted in what one moment of decision making may involve, represented in Figure 17 below:
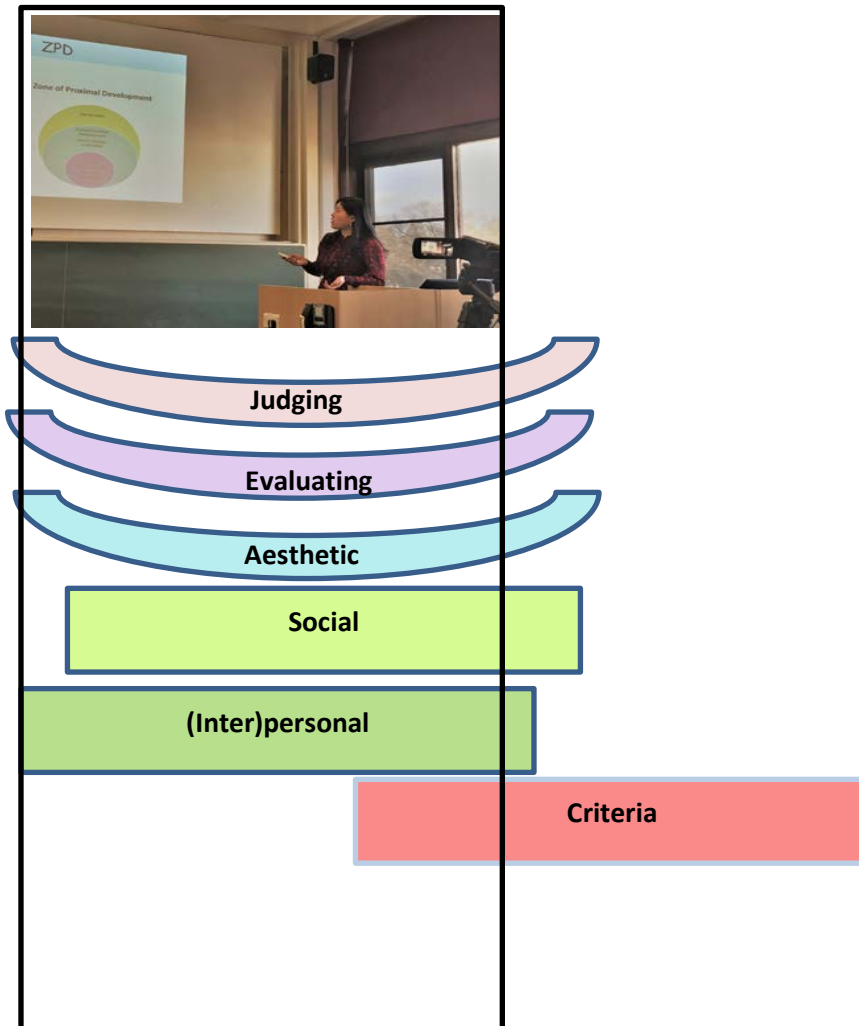
Figure 17: Rating Decision Making in Action

Particular lenses and filters are not being applied at this given moment (e.g. marking and technical lenses and ranking filter). Those lenses, and filters within the rectangular black box are being used in the decision-making process. The Criteria Filter is shaping the process to a small extent – only a third of the filter enters into the black box which represents the decision making. A vast amount of combinations are applied by the teacher assessors – the process represented in Figure 17 is a minute snapshot of the rating process with new combinations of tools and strategies being deployed and enacted continuously. Ascertaining what is happening within this black box of assessment decision making is complex. I wager that in the universe of assessment decision

making much dark matter pervades this process remaining undetected and invisible to researchers and those engaged in the decision making alike.

Adam, Daniel, Georgina and Tracey look for, and evaluate the relevance of, a complex array of criteria when they assess AOPs. It has been demonstrated that teacher assessors appraise how the descriptors fit the context and apply to the task and performance. The social learning environment is the unit for analysis (Moss, 2003) showing that a key factor which shapes the assessment construct is the social environment and the social filter the assessors employ. In answer to RQ2, the assessors look for broader qualities of readiness (cognitive skills, communication skills, social and emotional skills and formative learning skills) to embark on academic study in the UK (see discussion on EAP construct in Section 8.4 below). They gather evidence on these qualities in the AOP performances and assessment opportunities external to the AOP assessment. The convention in language assessment of mapping a performance onto a scale is not happening in AOP assessment practice in the field sites (validity and fairness implications are considered in Section 8.6 below). This is due to the inability of criteria to account for situational factors, a strong interpretive element, and the teacher assessors' inclination to allow judgements of the person to factor into the discussion of the AOP performance. The main factor shaping the complexity of the processes is that the teachers lean towards making how the student functions in the whole social learning environment on the EAP module and programme the unit of analysis, rather than the AOP performance.

Chapter 7   Student Perspectives on AOP Assessments.

7.1   Introduction

In Chapter 6, teacher assessors' voices were represented on how they approach the assessment task on their modules, and now I give the floor to the student test takers. The analysis complements the previous two chapters, which revealed what teacher assessors in the two sites do, look for and value. In this chapter, I present what students believe to be important in AOP assessment and what lies behind decisions they made regarding AOP preparation, delivery and reception on their EAP modules. It is important to garner information from students in order to understand their approaches to the assessment. This user perspective has deep roots in assessment theories on face validity and washback. Students, as will be shown in the ensuing analysis, offer valuable insights which have ramifications for the validity and fairness of assessment practice. An assessment task may not be interpreted and implemented in the way assessment developers and teachers predict. Salter-Dvorak (2016b) evaluated a development in the curriculum which involved students fulfilling an oral presentation task as a means to improving a written piece of work. Salter-Dvorak found that the students focused on delivery aspects rather than argumentation development as had been intended.

In the following sections, I present an analysis mainly of the student interview data (see Section 3.8.3.3 on interviews). In FS1, I interviewed five students out of the 14 students in the group, and in FS2 I interviewed three students out of 11 in the class (see Section 3.7.1 for informant details). The interviews with students and lesson observations were targeted at understanding what factors shape the students' thinking and stated approaches to the delivery and reception of the AOP assessments on their year-long modules (addressing RQ3: What factors determine the EAP AOP assessment construct?). The extent to which the students adapt their practices to suit their individual goals and ways in which they may differ from the assessment construct as exemplified in the rubric is discussed in the ensuing analysis. This contributes to our understanding of student perspectives on the EAP AOP assessment construct (addressing RQ2: What is the EAP AOP assessment construct?). My presence in lessons and at the AOP assessment events, and field notes made throughout the research have also aided analysis by situating accounts and gauging the plausibility of students' accounts. The factors, which have come through from the data most distinctly, are the personal, content, language, assemblage, interaction, confidence, formative, and summative. I generated these core codes using NVivo node and memo writing features (see

Appendices K and L for examples of S2 analysis process). These factors have been separated so as to emphasize the role of diverse elements in AOP practice. They are inextricably connected and integrated in the processes engaged in during AOP tasks, however. It is hoped that the presentation of components in this chapter does not overemphasize the role of specific factors and deemphasize the integrated relationships between them.

7.2   The Personal Factor

The personal factor refers to how individual trajectories, goals and motivations influence thinking and actions. In the interviews before and after the presentation, the students were asked what the purpose of the presentation was for them as individuals. The students cited motivations such as an opportunity to improve linguistic and communicative competences, to develop understanding of the research topic and develop academic skills, to gain a good mark, to inform their peers and to practise public speaking:

> *Excerpt 1:*
>
> Ivan: I think I can't improve my English in such a short time, but I can concentrate on making a good content and at least will probably improve my English. (FS1, IBP with Ivan)
>
> *Excerpt 2:*
>
> Henry: When I do the research {for the AOP} I know something about this {process of devising, analysing and presenting questionnaire data}. (FS1, IBP with Henry)
>
> *Excerpt 3:*
>
> Anna: I want to get the high mark is the first one, hehe. And then give the students our group presentation a more deeper understanding and let us know what we do in our group presentation. (FS1, IBP with Anna)
>
> *Excerpt 4:*
>
> Tara: My purpose? Umm, express what I think about that topic and the practice for the mind, my mind for the speaking in front of many people. (FS2, IBP with Tara)
>
> Excerpt 5:
>
> Aimee: To improve speaking skills and to get used to speak at the audiences. And also I think there are many opportunities to presentation in the undergraduate so get used to prepare and demonstrate what I want to say is the main purpose I think. (FS2, IBP with Aimee)

The students' conceptualisations of the main individual purposes of their AOPs expanded and evolved. For instance in the interview after the presentation, Ivan said the main task purpose was to improve linguistic competence, content skills and his ability to work in a team. For Anna, the main goals cited in the interview after the presentation were to improve English speaking skills, study skills, ability to communicate content and to improve group work skills.

The motivations behind students' higher goals for the AOP task related to their overall reason for choosing to study in the UK, their individual English language exposure and educational background, how set they were on meeting certain preordained aims versus emergent aims, and their confidence and desire to shape the assessment task for their own purposes. The next two sections demonstrate the importance of understanding students' diverse trajectories and degree of agency they exert over their actions when endeavouring to build a nuanced picture of their approaches to the AOP tasks.

7.2.1 Individual Diverse Trajectories

The students had different paths to the IFP and Pre-master's programmes. Their previous education, dispositions, chosen disciplines, hobbies, and family and friendships are some of the areas which have an influence on the choices they make. I illustrate such influences by discussing three students' thoughts on improving their English. Aimee, Tara and Marina, students in FS2, had resided in Japan for most of their lives and all identified themselves as Japanese. They had very different life experiences, including various exposure to and use of English.

Excerpt 1:

Aimee: Yeah so uh why I came here? Because I really want to speak English better with my childhood friends who can speak English better. They still live in America. (FS2, IBP with Aimee)

Excerpt 2:

Tara: I've been {to the UK} before but maybe for a week with my parents. So I didn't study abroad before. {…} I have a less chance to talk in English {before the IFP}. About writing I'm not good at but the teacher told us how to write the essay well so it's ok but speaking is difficult to me to express what I think.

(FS2, IBP with Tara)

Excerpt 3:

Marina: So I talk to my Dad in English and my mum in Japanese. And I used to live in England for 2 years when I was about 10 and that's why I can speak a bit of English, hehe! {…} So I chose the international foundation programme because, the main points because I really wanted to improve my English. We can study English in Japan but I think it will be better for me, I can get the knowledge more and improve my skills more in the country where English is spoken.

(FS2, IBP with Marina)

Aimee came to study in the UK in order to improve her spoken English so she can communicate in English better with her childhood friends, who live in America. Tara had visited the UK once on holiday before going there to study. Tara thought her spoken English was not good and deeply wished to improve. She gently expressed, throughout our interviews and conversations, a haunting frustration that she takes a long time to formulate what she is saying and cannot fully

express herself. She wished to be more confident in using English. The quest for achieving that inner peace and to be content in being quiet as well as vocal is something I have found is often lost in the discourse on oral communication. Marina lived in the UK for a short while when aged 10 and continues to speak to her father in English. She chose to study in the UK to improve her English, but feels it is at a good level to communicate already. Her teacher Georgina referred to her as 'practically a native speaker' but Tracey argued that her exposure to English and proficiency was not quite on the level of what she would describe as 'a native speaker of English'. Friendship, communication skills development and character development -- profound human concerns -- provided the momentum behind Tara and Aimee's thinking and actions behind choosing to study on an IFP at a British university.

These accounts show that the students have various motivations for enrolling on EAP programmes. This reveals the very reason why many EAP courses involve the development of a complex make-up of literacies and competencies. These 'international students' bring with them diverse linguistic backgrounds and repertoires. Aimee, Tara and Marina, who all identified as 'Japanese', highlight that definitions and distinctions such as 'international/home' 'native English speaker/non-native English speaker' are fuzzy concepts. It is difficult to determine whether Marina is classed as a second language user or first language user of English. At best, people given such a label as international student or second language user could be connected by some overlapping similarities but have no one feature in common between particular individuals – an example of Wittgenstein's concept of family resemblance. There is also overlap between the characteristics that a home and international student and a first language and second language user of English may share.

The students' personal trajectories initially shape their beliefs as to their personal strengths and areas to improve. Language development is a driving concern for enrolling on the IFP and for the students in FS1 enrolling on the Pre-Master's. However, the nature of the course and tasks shift the level of emphasis on original goals and their goal setting behaviours. Aimee, Tara and Marina regarded the IFP as an opportunity to develop their competence in their English language proficiency, but their accounts on the role of language development in the AOP tasks vary and do not always align with their motivations (see discussion in Section 7.4 below).

The degree to which the students shifted their focus depended on their level of commitment and belief in their initial personal goals, and the ability and usefulness of allowing their goals to shape AOP assessment practices.

7.2.2   Student Agency

When speaking to students, some questions were met with long pauses and puzzlement. Naturally, as with the teacher assessors, motivations for their practice were sometimes not visible, accessible at the time of interview or multi-faceted so difficult to crystallise into a few sentences. However, I was struck by how forthright the students were at times in their convictions which drove their approaches to the assessment. It could be assumed that in an assessment context in which teacher assessors evaluate the performance that the students would have limited space for manoeuvre and would therefore conform to expectations at a social level. However, this was not the case, with some students engaging in a high degree of 'individual positioning' in which they make minimal change to behaviours in order to align with conventional expectation (Widdowson, 2020, pp.90-1).

Ben's determination in fulfilling his overall goal of developing his ability to speak English spontaneously was palpable in the interviews and detectable in the AOP performance itself. Ben spoke about how his peers were preparing scripts of their monologues, and had asked him pointedly why he did not have a script. Ben's justification was that he wished to use the AOP as an extemporaneous speaking act. Although on reflection he said he would prepare his speech to a greater degree in order to improve the delivery in his next presentation, he was steadfast in his reasoning behind not having prepared a script on this occasion. Ben recognised that his delivery would have been smoother if he had prepared a script, however Ben's ultimate aim for being on the programme was to speak spontaneously and this is what he would do.

> Yeah yeah my aim is speaking naturally. So at that time I is improvising and other people is using a script and repeating at lunchtime. My South Korea co-worker repeats a script. But another of my South Korea friends point out 'your script, where is that?' No I only improvising.
>
> (FS1, IAP with Ben)

Ben shared in the interview that in his 26 years' learning English and in English classes at school in South Korea there was little opportunity to develop his spoken English. Ben divulged that he wished to speak English very well in order to communicate with people around the world including in the UK. The Pre-Master's programme was his time to immerse himself in an English-speaking environment and develop his spoken English language abilities. This was his desire from the outset and Ben exerts his agency to use the assessment task for his own specific purpose. On the other hand, for Ivan, his main personal goal was to improve his spoken English, but the multimodal nature of the task and importance of obtaining and reporting data effectively result in him switching attention to content and research skills. Ivan believed this is where the most improvement could be made and his energies were therefore diverted to the content skills

involved in the AOP task. (See further discussion on the influence of the content factor in Section 7.3 and the summative factor in Section 7.9 below).

Accessing students' motivations, on the one hand, points to cases of students adapting their original, personal goals to various degrees to align with the AOP requirements and features. Conversely, Ben did not seem to alter his behaviours in line with the AOP task. Ben's desire to improve his spontaneous speaking skills dictated that the AOP for him was an extemporaneous speech act. The students when approaching the task in their actions were asking or neglecting to ask: How can I shape the task to work for my goals? How are my goals compatible with the task at hand and the teachers' task representations?

The students had diverse goals associated with their particular use of the AOP, which had the potential to greatly shape their AOP delivery. Ben was primarily focused on developing spoken language abilities. Other students such as Ivan redirected focus to developing content and research skills; developing language was a by-product of the AOP task. However, it was the very specific motivation, for example Ben's focus on improving his extemporaneous speech performances, not a general focus on language, which shaped his decisions and actions. Accessing the individual students' motivations enhances our ability to make sense of the individual performances (see further discussion on seeking student perspectives in Sections 8.6.4 and 9.4). We can see the performances and assessments in a new light when considering the linguistic and non-linguistic factors which feed into their thinking.

This section has included examples which suggest that developing language is a key driver influencing student decision making in studying in English medium, but from delving deeper into students' thinking and behaviour this is not strictly the case. Some students on the EAP modules experience a re-direction of focus which is discussed next.

7.3   The Content Factor

Students in interviews stated the improvement of English as a core reason for completing their Pre-Master's programme and IFP at UK institutions. However, in the AOP task the content, rather than language use (see role of language in Section 7.4 below), was revealed by the majority of students to be their overwhelming priority.

This focus on content is most clearly depicted in FS2 when students were asked in individual interviews to consider which elements they believed should be assessed in their IFP presentation and what weighting should be assigned. During the interviews in FS2, I drew a circle on a piece of paper and asked students to divide the circle into segments to make a pie chart which reflected

their suggested criteria and weighting for the summative AOP assessment given their experience of the task (see Section 3.8.3.3 for methodological account of interviews). Marina, Tara and Aimee allotted the highest weighting to content:

> *Excerpt 1:*
>
> Content is 50%. … I think presentation skills is 35 to 40ish and the rest is language.
>
> (FS2, IAP with Marina)

> *Excerpt 2:*
>
> Interviewer: So you've written {on a pie chart on paper} content 50%, language 20%, and presentation skills 30%.
>
> (FS2, IAP with Tara)

> *Excerpt 3:*
>
> So this course's aim is to improve language so language might be more higher. I would put for a presentation assessment on an IFP, I think 80% is a bit heavy, so 70% for content and 25% language and 5% presentation skills. I think IELTS is 80% on language and 20% on content!
>
> (FS2, IAP with Aimee)

The students had been given access to the rating scale, however their responses to what should be assessed diverge from this. Content receives 50% of the weighting from Marina and Tara and 70% from Aimee. Even though Aimee expressed the belief that the aim of the course was to improve language, she dedicated a modest 25% of the weighting to language. Aimee makes a telling comparison of her perceptions as to how language and content are weighted on the IFP and in IELTS: 'IELTS is 80% on language and 20% on content'. It indicates one such motivation behind emphasising the role of content is due to a gearshift some students suggest they experience when transitioning from general English classes and large-scale standardized English language tests to EAP learning and assessment on their programmes.

7.3.1   The Gearshift in EAP

The gearshift refers to students manoeuvring differently to change their direction of travel and focus. Students, on their modules, experience shifts away from focus on language use to content.

Aimee, a student in FS2, shared how she felt her EAP classes primarily focus on developing abilities related to content. Aimee argued that the areas which demand greater effort and attention relate to the mastery of academic skills. Some of the students had transitioned from completing English tests (at school and English proficiency tests such as IELTS) which did not

require research to be conducted and the selection and command of new and complex academic content knowledge and conventions.

Students have diverse trajectories leading to their time on their EAP courses (as evidenced in Section 7.2 above) and for some students, such as Henry in FS1, this gearshift was not necessary. Henry had delivered an academic presentation in English as part of his undergraduate studies in the USA. In contrast, Ivan in FS1 shared how he was daunted by the content aspect of the Pre-master's AOP because he had not delivered such complex and unfamiliar content in English before. He shifted to devoting energy to the content skills. This points to the need to consider individual differences regarding experience of presenting academic content in English.

The increase in emphasis on content demonstrates that the EAP assessments are viewed by students as more than language assessments. The students described their EAP presentations as a task of compiling and delivering academic content *through* English. While in IELTS and tests in their English classes at school (in the cases of Aimee and Ivan) content is used as a decorative framework to hold up the language, EAP assessments have a more intricate and balanced relationship between content and language. In the EAP AOP assessments in the two field sites, the students seemed to view language as the means by which to hold up the content and performance (see more discussion on the role of language in Section 8.5 below).

7.3.2   The Indicator of Academic Ability

When describing content in the AOPs the students emphasized how important it is for student presenters to display features of *academic* content in their AOP. The students spoke about the importance of 'analytical', 'accurate', 'logical', 'verifiable', and 'convincing' content. Marina valued a presenter who demonstrated understanding and a command of their content. Tara looked for how presenters were concise and discerning in their selection of information and how they pieced it all together. Aimee used the word coherence numerous times when she spoke about AOPs and when talking more about this revealed that she valued a well-constructed and focused narrative with an 'understandable opinion and accurate information':

> Aimee: The highest point is content, how easy to understand or follow and coherence. Understandable opinion and accurate information. And after that the language.
>
> (FS2, IAP with Aimee)

The students explicitly highlighted the importance of analytical content, stance and critical thinking in their academic presentations. They were getting to grips with concepts such as 'critical thinking' and 'coherence' and at times it was difficult to ascertain how they were conceptualising these terms. They stated that their energies were directed at developing clear organised content and line of thinking. The AOP for many students was primarily a showcase of their emerging

understandings and ideas on how academic content and conventions are manifested in an AOP. In the AOP tasks on their modules, the content-related aspects where deemed by the majority of students to be what makes the presentation sufficiently academic.

Tracey, the co-assessor in FS2, spoke about how she viewed her role as gauging students' 'academic nous' and many of the students have picked up on this function of the AOP. Having worked as an EAP practitioner, the ambivalence as to the emphasis on content resonates with me. The alignment with practices in the academy is important, and many an EAP practitioner has spoken about the need to un-train students from IELTS. On EAP programmes there is a conundrum as to the feedback given on content-related aspects as someone who is not a content specialist but is expected to give feedback on aspects such as argument development. EAP teaching and assessment is attempting to strike a workable balance but there are pitfalls with being in a liminal space. EAP practitioners' positioning in the liminal space means they are tasked with finding ways to construct workable and justifiable bridges to the academy. The students when emphasising the role of content are picking up on this discipline and target domain alignment (potentially the views shaped by the curriculum, materials, tasks and teachers' views) and the role that 'content' plays in indicating how they will cope in UK academic settings.

A further complication is discerning where content and language end and begin. Subject specific terminology may be required to communicate deep understanding of ideas in a particular discipline. The students in the fieldsites spoke about potentially using more 'complex' words in their AOP tasks, but the language was not seen as a particular indicator of academic prowess. I lay out ways in which students approached English language development in the next section.

7.4 **The Language Factor**

Student perspectives on the role of language in the two AOP tasks at the different sites are complex and nuanced. Although all of the students wished to improve their English on the course through completing the AOP assessment in both field sites, there was varying emphasis on the importance and role of language in the AOP performance. While Ben was wholly motivated by improving his spontaneous English speaking skills in the AOP tasks, Anna, after deliberation and discussion, stated that 80% of the AOP purpose was to improve her English speaking skills.

> Anna: Maybe 80 percent to improve my English and 20 percent academic learning.
>
> (FS1, IAP with Anna)

The remaining students across field sites when asked what role language plays in the AOP expressed strong beliefs that language was not so significant in AOPs. I explore their views on this topic in the next section.

7.4.1   The Minimal Role of 'Language'

Marina and Aimee in FS2, and Fiona in FS1 from initial analysis of the interview data seemed to significantly downplay the importance of language in delivering an AOP. Aimee stated that 'language it not matter'; Fiona said 'language is just part of the presentation' and Marina argued that she 'doesn't really look at language' believing that it was not an area that should be assessed in an AOP task, even on an EAP module.

Excerpt 1:

Interviewer: And you're also assessed on language. So did you notice anything about people's use of language at all in the presentations? Is there something you thought about [or?]

Aimee: [I] didn't realise something about language.

Interviewer: So presentation skills you noticed their eye contact and body language. Content you noticed how narrow the topic is, the language?

Aimee: Language it not matter.   (FS2, IAP with Aimee)

Excerpt 2:

Marina: I won't really look at the language. I think the content is very important but also the presentation skills as well.

Interviewer: So you would delete language? Even on EAP courses? You would not include the language?

Marina: Yes. Yes.

Interviewer: What about if someone's level of English is quite low and they cannot express what they want to say? How does that fit in?

Marina: Eek! Even if their English isn't perfect, if we can understand what they're trying to say that's ok.   (FS2, IAP with Marina)

Excerpt 3:

Fiona: I won't be worried about language. Because language is just part of the presentation. We should concentrate more on what we're talking about. Not what kind of language we choose to talk. (FS1, IAP with Fiona)

Aimee and Marina expressed strong opinions about the importance of language in the presentation in these excerpts above. The students overall in FS1 and FS2, in comparison to content, place less importance on the role of language in the AOP. Further discussion with the students revealed a complex notion of what they were defining as language and what aspects of language use they prioritised in the multimodal AOP on their EAP modules, discussed in the two

following sections. What may have been seen as a demotion of the role of what Aimee and Marina are referring to as language initially may seem surprising because the students were studying on an EAP module and have both stated that a main aim of the course, module, being in the UK and their AOP assessment was to improve their spoken language skills. The seeming contradiction may be in some way explained by long-term goals not meeting the short-term requirements of the assessment task. This is elaborated on further in the next section.

7.4.2   Longitudinal Language Development

Students who had stated a motivation for being on their courses was to improve their spoken English indicated that this was not their focus in the AOP assessment. For instance, Ivan when undertaking the AOP task, communicated that improving English language abilities was not his focus short-term.

> Ivan: I think I can't improve my English in such a short time, but I can concentrate on making a good content and at least will probably improve my English. (FS1, IBP with Ivan)

The justification behind not focusing on improving and showcasing language ability for certain students was the recognition that developing language ability was something which they were developing over time and through doing multiple tasks. It was described as a crucial by-product of doing the tasks on their year-long programme. Ben is a notable exception to this as his AOP goal and overall goal of improving spontaneous spoken English were harmoniously unified, although deemed incompatible by the teacher assessors from Ben's AOP performance

When talking about their own performances, the students often referred to their use of language with a view to improving in the long-term.

> *Excerpt 1:*
>
> Fiona: For the language part I need more practice and more words to pick up.
>
> Interviewer: More of a range of vocabulary?
>
> Fiona: Yep. And more control of my speed because in the end I think I talk too fast.
>
> (FS1, IAP with Fiona)

Fiona stated that when watching presentations she focuses on communicating the content and is not too concerned with vocabulary choices and use of a range of language. However, the excerpt above shows that she notices these language features, as well as pace, when talking about what she would improve in her own performances.

These episodes and shifts in the way students talk about linguistic features reveals that students may have different standards at different points in time when conceptualising the role of

language use for the short-term AOP communicative purposes and their notions of, overall long-term communicative competence. Language may not have a strong summative function, but be valued for formative purposes (see analysis on this factor in Sections 7.8 and 7.9, and discussion in Section 8.5). This may be explained by the fact that language and content become inseparable in the communicative performance. The students' perspectives on 'content' and 'language' may be regarded as confirming the notion that content and language are two sides of the same coin (Gillet, 2013). The content and language become merged together when considering preparation and delivery of the given assessed AOP communicative act. This entails that language is viewed as a medium and not the target.

7.4.3   Spoken Language as Vehicle

Establishing which linguistic features the students refer to when using the word 'language' develops a closer understanding of their orientations towards language use in the AOP assessments. The type of communication viewed as characteristic of the AOP is what shapes the perceptions on what language use is deemed appropriate. When 'language' is viewed as unimportant it tends to refer to verbal communication and primarily grammatical features. This perhaps explains the conviction with which the students downplay its relevance. It is however, crucial to understand orientations to language use in AOPs further. Spoken language used in the AOPs does not need to adhere to a particular standard according to some of the students. Marina and Aimee focus on the ability of the presenter to convey the main content.

> Marina: Even if their English isn't perfect, if we can understand what they're trying to say that's ok.
>
> (FS2, IAP with Marina)

> Aimee: So if the grammar was wrong, it not matter. I think it is not much big issue for the presentation think. It's not big issue I think.
>
> (FS2, IAP with Aimee)

These students when expressing opinions on the unimportance of accurate grammar in speech tend to refer to what has been argued in ELF(A) research to be a legitimate non-standard feature of a speaker's repertoire in ELF interactions. Marina refers to the types of grammatical mistakes she makes such as subject-verb agreement and gives an example zero-marking third person s.

Marina: I do a lot of mistakes.

Interviewer: What kind of ones?

Marina: So S V

Interviewer: Oh subject verb agreement

Marina: Yeah I really can't do that. I don't think it's important in a presentation…

(FS2, IAP with Marina)

The students argued that many 'inaccuracies' in verbal communication are not noticed unless there is a breakdown in communication. This is in line with research conducted which has demonstrated that third-person s is communicatively redundant (Seidelhofer, 2005; Breiteneder, 2005). It can be inferred from Marina's comment about the unimportance of subject-verb agreement in a presentation that she does not view Standard English as the desired benchmark in all communication, including in assessments in English. This shows that in an assessment setting many speakers using English in lingua franca settings are viewing the language as a 'vehicular language' (Mauranen, 2006, p.124). Marina does not pay attention to features which have low communicative valency. The students do not seem concerned by potential impact of certain non-standard usage of English in the AOP task. This is in some way explained by the students' lack of emphasis on the importance of spoken language in the AOP. From these glimpses into the these students' thinking, the students seem to be interested in the ends and not in the means when it comes to describing the role of spoken language in AOPs.

Why is verbal communication deprioritized by these students in what is often referred to in EAP circles as a 'speaking assessment'? The students' recognition of the multimodal nature of communication in the AOP goes some way to understanding the minimal concern with verbal communication and especially grammatical accuracy in spoken communication in their assessed AOPs. These attitudes to the role of linguistic features in AOPs are elaborated on next.

7.5   The Assemblage Factor

Assemblage refers to the grouping together of diverse resources or elements. It is a term used in different domains, notably in art. In the realm of art 'assemblage' describes the creation of three-dimensional art forms made from various disparate materials: it could be understood, rather crudely perhaps, as a three-dimensional collage. A piece of written work such as an essay may be conceptualised as a collage and the AOP an assemblage. The slide contents, speech and the presenter's limbs and facial features are brought together especially in a three dimensional performance. The resources at the students' disposal are linguistic and non-linguistic.

The students viewed the spoken monologue as a small part of a presenter's kit of materials. They consider the assessment from a communicative frame and have a sense of what modes allow presenters to communicate meaning. There is an acknowledgement of the linguistic features: the spoken utterances and the written text on slides; as well as the non-linguistic features: graphics on slides, the use of eye contact and body language.

7.5.1   Linguistic Modes

Aimee and Marina considered the multimodal make-up of AOP communication, with the spoken and written modes by which a presenter is able to relay information to their audience.

> *Excerpt 1:*
>
> Aimee: The main message they many times saying and the main message in the presentation {slides} and the speakers emphasize the main opinion in the presentation.
>
> > (FS2, IAP with Aimee)

> *Excerpt 2:*
>
> Marina: So the PowerPoint will show what they want to tell as well I think. I think we can understand what they want to say.
> > (FS2, IAP with Marina)

The role of 'language' in AOP communication, through the dialogue in the interviews, was shown to be more important to the students than at first depicted (as discussed in Section 7.4.3). The permanence of the writing on the slides was a factor which meant the writing should be as accurate as possible for Marina. When asked about any instances when accuracy is important, Marina acknowledges the role of written text on slides:

> Marina: Spelling mistakes, because all the audience look at the PowerPoint these are mistakes that we would all notice a lot so I think it's important on the PowerPoint.
> > (FS2, IAP with Marina)

According to Marina, while the audience may not catch verbal inaccuracies they would notice any mistakes in the text that is displayed for longer and may be accessed after the presentation. Aimee shared that one of her bug bears was font on slides that was not easy to read. This was distracting for her as she often focused on the text more than the spoken utterances of the presenters. Many of the students spoke about the need for effective signposting language and accurate pronunciation of key words if these were not on the slides, recognising the relationship between the spoken and written modes for successful communication. The emphasis placed on the written language in the visual aids explains why the spoken utterances were viewed as less important for some students.

The use of visuals on slides, body language and eye contact that the AOP affords was highly valued by students in AOP communication.

7.5.2 Non-Linguistic Modes

The importance of eye contact, body language and images and graphics on slides was emphasised by the students. Henry in FS1 and Marina in FS2 believed aspects of non-verbal communication were extremely influential in delivering an effective presentation.

> *Excerpt 1:*
>
> Henry: So the first important thing is to stand face to the audience and your eyes need to look around. This is very important. Probably like your hand have a little bit of action and when you show a graph probably use shaped hand.   (FS1, IAP with Henry)
>
> *Excerpt 2:*
>
> Marina: Yeah body language is like EVerything. They {Eye contact and body language} are a MAssive thing. I think it's really important. I was reading from the paper today but I think if I do that the audience will have less interest in my presentation as well and it's easier for them to understand if I use body language and look at them. I think it's really important.
>
> (FS2, IBP with Marina)

The students' views are, to some extent, in line with research conducted into multimodality, posthumanism and poststructuralism (as outlined in Sections 2.8.4-2.8.5) which argues that communicative proficiency is 'shaped by artefacts' (Kuby, 2017) and requires that communicators attend to the 'full range of communicative forms' available (Jewitt, 2017). The students recognize the PowerPoint slides are a powerful artefact shaping delivery and features of non-verbal communication as effective communicative forms. Meaning arises from multiple semiotic resources deployed by the presenter, including non-linguistic modes. Many oral presentations given in academia may not use visual aids and verbal communication is the sole delivery device. However, at the two fieldsites, on the EAP modules, the students are encouraged to use supporting slides. The students devote a large part of their attention to the slides, which explains why spoken language is viewed as not so important.

Reliance on notes was an aspect which students considered to be an important topic when evaluating qualities of an effective presentation. Most students divulged their desire to move away from the use of scripts, even if in the final AOP they had a script to hand -- in some cases in their hands. They wished to have the ability to be more flexible in their use of non-verbal communication. The students viewed the use of notes as a barrier to effective eye contact and body language which they viewed as channels of contact with their audience and a crucial part of audience engagement. All students attested to notion of the AOP as an interactive performance

and the importance of the aesthetic and emotive qualities of their performance. There was a longing among the student informants to engage the audience in order to deliver their message and entertain.

7.6 The Engagement Factor

The interactive nature of the AOP shaped the students' AOP delivery and reception of their peers' AOP performances. An 'interesting' presentation which 'catches attention' was highly sought after by students. The overall priority for most students is communicating the main message, but they greatly appreciated aesthetically pleasing slides and emotive delivery.

There were instances in the field when students became animated when talking of the bond that was created during the AOP, their preoccupation with the audience's reaction when presenting and their emotional and aesthetic preferences which heightened their engagement as audience members.

7.6.1 Creating a Bond

The student informants in both sites expressed a desire to deliver a presentation which connected the whole group. In FS1, this was a feature of presentations that the teacher Adam valued. It seems to have been fostered in a number of his students through the course and assessment design.

The students in FS1 and FS2 were concerned with not only being *engaging* but being *engaged* with the audience. Aimee thought that being engaged with the audience -- creating a connection -- was a vital part of communicating in the AOP.

> Interviewer: What does it mean if someone is engaged with the audience?
>
> Aimee: They make the audience interested in the topic or participate in the presentation.
>
> Interviewer: Like your Kahoot.
>
> (FS2, IAP with Aimee)

In her assessed presentation, Aimee used an interactive game using an app called Kahoot. The audience guessed the answer to a question related to her presentation content.

Fiona in FS1 talked about how a presenter is effective if they show investment and create a bond with the audience.

> You will have eye contact with each other and maybe you can use the examples that we have in our daily life that they are familiar to so they will have the same feeling as you are.

The students were seeking ways to create a positive and dynamic interaction with the audience. Fiona sought ways to create a bond in the preparation stages in her slide design and when practising verbal and non-verbal communication. She also was keenly aware of the audience's response while she was delivering her assessed presentation, which is discussed below.

7.6.2   Gauging Audience Engagement

Both Fiona and Henry talked about wanting a connection with the audience and to have fun. They shared in the interviews how heightened their awareness is to the audience members' reactions when presenting and watching from the audience. Henry gauges the audience's interest by looking at their facial expressions.

> Henry: Yeah I think during the presentation I do like to stand face to the audience and actually I do look around and see their eyes because I want to know what the audience think about that.
>
> (FS1, IAP with Henry)

Fiona, like Henry, recounts how she consciously gauged the audience reaction and tried to change performance accordingly. She narrated an episode during her assessed presentation when her friends were talking and believes she deciphered what they were saying while she was presenting. She divulged that she also noticed when Adam and Daniel were looking at their paper while she was presenting. Racing through her mind is not only her speech but her concerns and perceptions regarding what the audience is doing or thinking. In addition to this, Fiona recalled how she is assessing her own performance as she speaks and making adjustments in line with what she feels the audience would appreciate. Multiple layers of voices assess Fiona's performance as she stands centre stage. The performance is not delivered as prepared, it is highly performative in nature (concept discussed in Section 2.8.3) and 'assembled in situ' (Canagarajah, 2018). There is a heavy burden of impression management (Goffmann) for the student presenters. (see discussion on The Affective Factor in Section 7.7 below). Perhaps, Fiona is putting too much pressure on herself in wishing for a (visible) positive reaction from her peers and teacher assessors who engage differently according to their aesthetic sensibilities.

7.6.1   Appreciating Performance Aesthetics

Instead of fixating on scores, in FS1 there was the strong desire to give an interesting presentation which would engage and garner a positive emotional reaction from peers and teachers. The students in FS1 and FS2 noticed when slide design was creative and pleasing to them and when the atmosphere that was created by the presenter was in some way inclusive and created a connection and bridge between the audience and the presenter. This depended on the aesthetic

sensibilities of the individual audience members. For instance, while teachers described Ben's presentation style as 'odd', Ivan praised Ben's presentation technique as it created a harmonious atmosphere:

> I think Ben did really good because his presentation is full of emotion, and he can make the atmosphere harmoniously maybe. (FS1, IAP with Ivan)

Ivan felt included and engaged in Ben's emotive delivery. This is a telling episode as Ben's presentation style appealed to a student in the group, however the teachers and I felt the presentation was, at times, difficult to watch. In field notes, I noted down that the pauses, uncomfortably loud speech, unfinished utterances, and difficulty accessing Ben's train of thought and purpose of his contribution made this a weak performance. The teachers described an 'odd' presentation style and difficulty accessing the main ideas. The atmosphere created is shown to be interpreted differently by the teacher assessors and the student Ivan in this instance. What is determined to be effective and appropriate interaction and pleasing communication styles is highly subjective.

The affective state of the presenter and audience has been touched upon in this section (7.6 on The Engagement Factor) in relation to the presenter's preoccupation with impression management and in the audience's emotional, aesthetic response to the presentation. The affective dimension is elaborated on further by focusing on the concepts and roles of confidence and motivation as described by the students in FS1 and FS2.

7.7 The Affective Factor

> Aimee: Um. Hmm. I think essay is think more slowly and more comfortable to think. But I think presentation is need more confidence and have a pressure for the present so it is stressful.
> (FS2, IAP with Aimee)

Aimee communicates that delivering a presentation is more stressful than other genres, such as an essay, due to the real-time face-threatening nature of the communicative act. The emotional aspects which affect the students' thinking regarding the AOP include concepts such as emotion, motivation, self-confidence and anxiety. The students gave narrative accounts of the difficult but essential task of controlling nerves and approaching the AOP task in such a way to convey positive affective displays of confidence and motivation.

7.7.1 Displays of Confidence

The students, except for Fiona in FS1, stated that they were nervous about the AOP. Fiona, although stating she was not nervous was conscious of the audience's reaction (as discussed in Section 7.6.2) and from my time with her seemed to be agitated by various aspects of the AOP.

Nerves is an aspect which the students feared would negatively shape their AOP delivery and reception. The students described how having notecards and scripts to hand were used to alleviate the fear of forgetting everything. There was also a dread associated with not being able to answer a question and being confronted with a sea of disinterested audience members. Communicating with confidence and conviction was what they desperately craved.

While confidence is a nebulous notion, the students and teachers (discussed in Section 6.7.3.3) in both sites described confidence and how this may be projected in various forms. Students recounted displays of confidence in the presentations which included having the courage to inject emotion into the performance, demonstrating a command of the content, using non-verbal communication with purpose, and speaking in an assured way without too much hesitation or a wavering voice. Despite the vagueness of what confidence constitutes in some episodes, the students and teachers have shown it manifests itself in many guises – like the concept beauty, it is found in diverse places and manifestations according to the eye of the beholder.

Ivan highlights how emotion, confidence and language are all intertwined in the multimodal AOP performance:

> Ivan: To me I think I could only use just simple words but maybe the markers want you to use some complex words and make it more academic maybe. And the fluency and pronunciation and your emotion maybe. Can you get your audience concentrate for you, or are they not interested in what you're saying.
>
> Interviewer: So with the emotion, how do you think that you do that? How do you show that?
>
> Ivan: I think maybe I was quite shy before, but yes one thing we have done is a presentation before in the listening and speaking class. And I think by this presentation I could find myself relaxed and free when I was presenting in the front. I think maybe I'm not very shy and I can be confident.
>
> (FS1, IBP with Ivan)

Ivan linked his confidence with the ability to command his audience's attention and attributed his growth in confidence to practice presentations in class. In his AOP tasks, Ivan performs as a new self. With the ability to relax, the freedom to express himself more freely is unleashed. Ivan crystalizes the elements that the students are mastering in their AOP performance: the technical features involved in achieving fluency and pronunciation, but also the emotional and affective dimension which shapes the production and perception of such features.

Aimee attributed the art of a comfortable viewer experience as one in which students' facial expressions do not distract and create a positive atmosphere.

> Aimee: Yes and their facial expressions, the smile is better than other face.
>
> Interviewer: What if the person's content is good but they have a serious face.
>
> Aimee: I would not mark them down but someone who smiles may get more marks if compare. It depends on the topic they are talking. The most highest weight is put on content but we should be conscious of the language and the performance. I think smiling is important for me. Because if the speaker's face is upset or angry I think what's wrong with them. If smile and look at me and if body language and to be confident I think attract and convince people.
>
> (FS2, IAP with Aimee)

A perceived lack of confidence, according to Aimee, diverts attention from the presentation content to an internal monologue focused on exploring the presenter's emotional state. It is interesting in this exchange how Aimee expresses what she would do as an assessor, not as a student. Aimee, who had previously stated that 'language not matter' is now considering its role in creating the performance aesthetic. Facial expressions is one such non-linguistic feature which creates an impression of the presenter who Aimee perceives as a performer. Aimee also talks about aspects involved in rating an AOP performance – a student may not lose marks for an unengaging performance but someone who delivers with energy and enthusiasm may gain an extra mark. Her call for being conscious of language and performance suggests that she believes these are somewhat powerful parts that shape the audience's ability to focus on the content.

What is considered appropriate and engaging body language was not discussed as an issue of equity and fairness by students in my time with them. The emphasis on particular use of non-verbal communication in presentations is regarded by Georgina, the main teacher in FS2, as potentially highly discriminatory. According to Georgina, non-verbal features should not be assessed in the gateway AOP assessments. I echo Georgina's caution; assessing posture, expressions, mannerisms and demeanour raises major concerns of fairness and equity if judgements on these contribute to the score on a gateway assessment.

Confidence was not a concept assigned purely to the delivery style. Marina communicated the need to be confident with content. In their practice presentation, Marina, Tara and Aimee (student informants in FS2) used a script. This action was taken in order to feel confident with their content and their ability to speak for 15 minutes. Without having internalised their content, they experience a lack of confidence in the delivery.

Marina: And then, I think I need to, I need to get more familiar about my topic and know more and then I'll have the confidence. Now I haven't got the confidence and that's why I'm reading from the script because I don't want to make mistake. So I have to get more knowledge about my topic and get the confidence. And then I think I'll get better with more eye contact.

(FS2, IBP with Marina)

The confidence with the content is what underpins the performance, allowing students to free up cognitive and physical energy to improve the delivery. The need to afford students the time to develop such confidence with content is essential – should confident delivery be expected on a 6-week pre-sessional course for instance. As was the case with the teachers, a number of students do not view the confident delivery as a substitute for poor command of content when talking about what makes a successful assessed AOP on their modules.

7.7.2   Displays of motivation

Levels of motivation shown by presenters was an aspect a few students mentioned as integral to a successful AOP performance.

Fiona believed that a perceived interest in the topic of the AOP would encourage the audience to listen more intently to the AOP.

Fiona: Um and I think if you are interested in the subject your audience will also feel your passion and will listen to you. I think this time this presentation much more better than last time. Because we had content and we tried to analyse the data and we designed the questionnaire by ourselves. We know what we're focused on and I think each of the classmates actively participated in this presentation. I can see everyone just feels excited about the presentation and I can see they prepared. I know one student prepared for the whole week. (FS1, IAP with Fiona)

The students in FS1 when speaking about the preparation and delivery stages of the AOP performances mentioned the level of participation and effort that individuals made in their own group and other groups in their class. Fiona highlighted the need for equal participation and preparation and how unequal effort was visible in the final presentations. She also recounted how some students were known to be putting in the effort, while others were notorious for their lack of contribution. The amount of effort and participation was recognised as an issue of integrity by Fiona. Fiona praised participation, preparation and dedication to the project. She also complimented the assessment design as she attributed active participation to the fact that the students designed the questionnaires themselves. The students had shaped and participated in designing, conducting and reporting research and were invested in the project and process.

Marina, a student in FS2, also believed that effort is noticed and appreciated. For example, if students make spelling mistakes this is not acceptable, even if their first language uses a different script, as the students should make the effort to check the dictionary or use a spell checker. Furthermore, Marina argues that in the Q&A students should be assessed on the effort that they dedicate to answering the question rather than competence in answering it.

Excerpt 1

Marina: I use the dictionary a lot and they should use a dictionary. It is about the effort they make to look up a word.   (FS2, IAP with Marina)

Excerpt 2

Marina: Even if they can't understand what you're trying to say, if they are trying to answer it I think that's ok. It has to be near enough. It's the effort that they make to answer the question.

(FS2, IAP with Marina)

In their discussions, the students revealed that the role of the students in assessment is to dedicate effort to make improvement and to apply feedback from their teachers. Effort is a quality that the students look for when making a judgment on whether an AOP presenter and performance is impressive. They consider the individual presenter and the presentation and gauge not only the final product but the process the students have gone through on the course which has been shaped by the students' level of investment and effort in the learning and assessment activities. Teacher assessors also focused on motivation, participation, preparedness, improvement and effort (as evidenced in Section 6.7.3.4). In this respect, there are parallels with what students and teacher assessors look for and value in the assessed AOP performances and this links to the student perspectives on the formative dimension of the learning, teaching and assessment on the module.

7.8   The Formative Factor

The formative dimension of the assessment was at the forefront of students' actions. Some students when prompted agreed that they evidently sought a good mark but they tended not to focus on this. The students were thinking instead of how the task would stand them in good stead in the future and how the performance would be received by students as well as the teachers. For instance, Aimee and Tara both refer to skills they would like to develop from completing the

assessment task when asked what the purposes of the task is for them. These include developing language and presentation skills.

Excerpt 1:

Aimee: Improve speaking skills and practise doing a task which may do on undergraduate course. Just to get used to present in front of audiences and present own opinion. Also the speaking with the correct pronunciation or signposting and also I think to make coherence the content is difficult. (FS2, IBP with Aimee)

Excerpt 2:

Tara: Purpose? Express what I think about that topic and the practice for the mind, for the speaking in front of many people. (FS2, IBP with Tara)

The student informants shared their willingness to make improvement based on feedback given on previous presentations. Their role in assessment as implementing feedback on assessment tasks, not only to improve their scores but their skill base and confidence is considered.

Excerpt 1:

Marina: I think that because we did a practice presentation before we did it and it helped us all and we all improved. We got a lot of feedback on that and that was really useful.

(FS2, IAP with Marina)

Excerpt 2:

Tara: Maybe it {the presentation} was good but practice for this semester I just tried to do my best. But this time there were many feedbacks about last time presentation and I know many improvement points. Georgina and Tracey were very kind and they gave me a lot of feedbacks so I could understand what other data to put in and more explanation. So it was very useful feedback. (FS2, IAP with Tara)

These segments of data indicate that the formative and practice assessments, which the students completed in the field sites, were crucial in developing skills and understanding expectations. The students are attuned to the usefulness of specific feedback and the development of not only themselves but their peers. The student-informants in FS1 shared that they had learned new skills and received valuable feedback from other students; in FS2 this was not expressed by the 3 student informants who spoke only about the usefulness of teacher feedback even though they had received peer feedback on two practice AOPs. Peer support is a feature which could be emphasized to a greater degree.

The recognition of the importance of making progress was not associated purely with the practice tasks; the summative AOPs were also acknowledged as formative opportunities. The AOP tasks were presented as an iterative process of developing and displaying an array of skills which in turn develops the students' self-confidence in diverse forms.

Excerpt 1:

I think giving a presentation in English is not only practice your speaking but also your confidence, like giving audience some body language and make some eye contact with others I think it's important. And it not only improves our speaking but also gives more confidence in speaking English.   (FS1, IAP with Anna)

Excerpt 2:

Um one thing I want to say is the presentation is really beneficial to our confidence. Maybe I didn't confident before the presentation and I would feel a little nervous but through the presentation maybe I can be more confident. And just feel free and relaxed. And maybe next time if we could make a better preparation we could feel more confident. Because the practice makes perfect. Hehe hehe!   (FS1, IAP with Ivan)

The act of delivering AOPs throughout the course has the ability to develop the students' communicative, social and emotional, cognitive and formative learning skills (converging with teacher assessor perspectives on readiness in Section 6.7.3). Such formative experiences result in greater self-efficacy beliefs and herein lies the greatest worth of the AOP assessment processes for many of the students.

7.9   The Summative Factor

The high-stakes nature of these AOP assessments would indicate that these students would be highly driven by the scores they would attain. The judgements made on the presentation performances have the power to determine whether students make the transition to their disciplinary studies. The AOP assessment in FS1 had the potential to contribute to a summative portfolio of work and the presentation in FS2 formed part of the final mark. The formative focus was palpable (as demonstrated in the previous section) in both sites but the focus on the scores – the summative dimension -- also factored into decision making, even if not as strongly across the student informants as I had anticipated.

In my time in the field, students described strategies which were adopted because of the summative function of the AOP activity. The following excerpt from an interview with Ivan encapsulates how Ivan believed that all students best focus their energies on developing research skills. Ivan wagered that the students could make most improvement in the content criterion. An

216

improvement in content skills could neutralise the lower mark he expected students may receive for the language criterion:

> I think we maybe we can concentrate better in the content because we can't use English very well, but maybe we can have a very good content. We can do a lot of work and research and questionnaires. (FS1, IBP with Ivan)

Aimee's motivation for coming to study in the UK was to improve her spoken English. Tara also believes her spoken English is not so good and would like to improve this. However, similar to Ivan in FS1, Tara in FS2 stated that she may 'skill up' her speaking over time while the preparation of content can be achieved in a shorter time frame (as discussed in Section 7.4.2), and this is where she feels her energies are best directed in the AOP preparation. For Ivan, Tara and Aimee language is developed from doing tasks and therefore an important formative goal, while content is an area students actively focus on improving for summative and formative purposes.

The reality that the central AOP task formed part of a suite of summative assessment shaped the students' conceptualisations of the task. However, it was striking how minimal the impact of the summative dimension was on both these courses in some respects (as set out in Section 7.8 above). The teachers emphasized how the students' abilities were developing or had developed over the course of a semester or year-long module. In FS2, the summative aspect was more pronounced than in FS1. This could be attributed to the fact the AOP would definitely contribute to the final grade, whereas in FS1 the students had a second attempt in the coming semester. The portfolio design on a year-long course seems to gravitate the focus towards the formative dimension. The impact of the teachers' views on the students' approaches was not clearly discerned and would be useful to explore in further research.

7.10 Chapter Summary

This chapter has presented students' perspectives on the AOP assessment on their modules. The analysis demonstrates that students exert various degrees of agency over their practice. They come with diverse motivations and reasons for changing their goals and orientations towards the AOP task and purposes of their modules and programmes (in Section 7.2). The AOP was depicted by most students as a showcase of academic skills and ability to communicate ideas drawing on a range of communicative resources (in Section 7.3). The AOP is not considered to *showcase* spoken language abilities but can *develop* them; spoken English is one means by which to communicate and necessary but not the crux of the communication for 7 out of the 8 student informants (in Section 7.4). The spoken delivery was not regarded as very important, by the majority of students, which is a striking finding as these assessments are conceptualised as speaking assessments at EAP practitioner events such as BALEAP PIMs (ramifications of the student perspectives on

language are discussed in Sections 8.4 and 8.5, 9.2 and 9.4). The students, in various ways, view the holistic communication of the main message as of paramount importance and consider the role played by different modes, especially the use of PowerPoint slides, in ensuring successful communication in AOPs (in Section 7.5). The students prioritise and appreciate the interactive, performative, affective and emotional dimensions of the AOP communicative acts (in Sections 7.6 and 7.7). There is clearly a strong formative dimension to the students' approach to the assessment task as well as a summative one. These findings offer sophisticated and profound insights into AOP assessment practice and make a contribution to working understandings of the complexity of the EAP construct, the relationship between linguistic and non-linguistic features in AOP assessment, and strategies used to enhance validity and fairness in AOPs, all of which are addressed in Sections 8.4-8.6 below. In future research outputs, I would dedicate space to explore the integrated nature of interrelated factors and set out more comprehensively the differences, variations and nuances in the student perspectives.

The next chapter synthesizes and develops the analysis of the data from teachers, students and artefacts from the previous four analysis chapters into core interpretations and themes relating to principles in how AOP assessment is done and experienced. The penultimate chapter discusses the navigation of a notion of an AOP genre, oracy demands associated with AOP tasks, the construct in EAP AOP tasks, the role of language in EAP AOPs and culminates in an examination of the ramifications of the findings for how we conceptualise validity in classroom-based assessment contexts involving prolonged contact between teachers and students.

Chapter 8   Discussion

8.1   Introduction

In the previous 4 chapters I presented an analysis of data from S1 and S2. In this chapter, I discuss key findings which contribute to our understanding of AOP assessment practices in UK Higher Education.

In this research I have developed richer understandings as to conceptualisations of the 'academic oral presentation'. Kaur and Ali's (2018) critical review of genre analysis studies on AOPs concluded that AOPs had been neglected in genre analysis. AOPs have been investigated to some extent for their linguistic (Zareva, 2009, 2016) and visual features (Morrell, 2015; Zareva, 2011; Hill and Storey, 2003) and their worth as learning and socialization tools (Duff, 2010; Zappa-Hollmann, 2007). However, a sustained exploration into manifestations of AOPs and their functions when used as assessment tools is an area which has received little attention. In Section 8.2, I give a reflexive account of my experiences navigating and negotiating the term 'academic oral presentation'. This builds on, and departs from, conceptions of oral presentations put forward in previous research (Swales, 2019; Levrai and Bolster, 2015; Hyland, 2009; Ming, 2005).

The oracy demands placed on students and staff in HE in tasks with an oral dimension have not been sufficiently documented, especially in AOP tasks. Studies have shown that oracy can be part of the assessment process and product (Doherty et al., 2011; Heron, 2019). I consider what findings indicate about the oracy demands associated specifically with fulfilling AOP assessment tasks on subject modules and EAP modules in Section 8.3.

I then narrow the focus of this chapter to a discussion of EAP AOP practice. The findings of this project have developed theoretical understanding of AOPs as EAP practice. In doing so, the findings fill a gap in theory-building related to EAP assessment and contribute to addressing the issue of 'the construct of EAP in EAP assessment being under-defined and under-theorized' (Hamp-Lyons and Schmitt, 2015, p.3). I outline and explore the complex features of the EAP AOP assessment construct (Section 8.4). I highlight that the complexity lies in the use of AOP assessments as windows into a wide range of skills which indicate students' 'ability to cope' in their future academic studies. Following this, the findings related specifically to the role of language in EAP AOP assessment are discussed, particularly in light of the range of communicative resources the AOP affords and of which students and teachers take advantage (Section 8.5). The project has generated specific validity concerns when using AOP tasks as language assessments. The findings enhance dialogue among scholars (e.g. Jenkins and Leung, 2019; Shohamy, 2018;

McNamara and Harding, 2018; Hall, 2014) on the role and construct of language in language assessments.

This investigation of oral presentation assessments offers insights into tools and strategies available to teacher assessors and students. User perspectives and approaches to the assessment have ramifications for validity claims. In light of the project findings, I extend conversations (in Section 8.6) on the implementation of naturalistic strategies in enhancing assessment validity (Lincoln and Guba, 1985; Moss, 2003) to high-stakes gateway AOP performance assessments in educational settings with prolonged contact between teachers and students.

8.2   The AOP Construct

The words academic oral presentation may conjure up an image of a human standing up, or sitting down, in front of a group of fellow humans with a concerted effort to emanate a relaxed demeanour and poise while speaking and gesturing eloquently and engagingly on a topic with some prompting from a prepared paper or a functional and aesthetically pleasing PowerPoint. This description may seem inviting, mechanical, or objectionable even to those who dislike how PowerPoint slides have become synonymous with presentations. It may be a distant reality from what first springs to mind when conceptualising what an AOP is in your mind's eye and starkly different to the last AOP you encountered. Perhaps the very notion of an AOP is instructive, or unhelpful for individual HE practitioners.

The response to requests for data on academic oral presentation tasks in this study has demonstrated that the term academic oral presentation proved to be sufficiently recognizable as a format and task in the academic settings which feature in this research. There was, at least from my perception, a reasonable degree of shared understanding between myself and respondents when I was attempting to limit the scope of the research to a focus on academic oral presentation assessment tasks and practices, bar a couple of instances of negotiating small differences between AOPs, poster presentations and vivas. I endeavoured to understand what this shared understanding may be of AOPs among HE practitioners. An AOP seems to involve humans delivering a prolonged monologue which transfers information, which displays academic knowledge and skills or contributes to an academic debate, to an audience within a particular time frame in an organized setting, with the opportunity for advance preparation and often the ability to support the speech with the use of non-verbal communication and multimedia. My accounts cohere with Ming's (2005) description that AOPs are a 'partly spoken and partly visual form of communication' (p.118). However, Ming's description does not distinguish the AOP from many other genres of communication which use visual prompts, such as poster presentations and

seminar interactions. The visual affordances in AOPs may be exploited by a presenter or viewed as an unnecessary embellishment. Reading a prepared paper or script with little gestural movement may be the presentation style of choice or one dictated by anxiety. The 'prepared' (Barret and Liu, 2016) and 'monologic' interaction (Swales 2019; Barret and Liu, 2016; Zareva, 2009; Hyland, 2009) are the most steadfast of features encountered in the AOP tasks I have been exposed to during this research. The core commonalities, or 'privileged properties' (see Section 2.6 above) are restricted; other than these characteristics the AOPs took on various guises. This representation facilitated communication about AOPs at a broad abstract level, but data on the make-up of AOP tasks and practices related to their design and implementation, reveal problems with any fixed conceptualisation of AOP assessment practice in Higher Education

When analysing examples of AOP assessment tasks in UK HE, fractures, reconfigurations and extensions continuously emerged in my mental imaginings of AOP tasks. The data from study one and two has demonstrated that the AOP is used as a learning and pedagogic tool as well as an assessment tool (as shown in Sections 4.3.1, 6.7 and 7.8 above). This supports findings of studies which have observed the learning gains of students when using AOPs on educational courses, from development of language skills and socialisation into disciplines (Sundrarajun, 2007; Duff, 2010). The formative dimension of assessment is palpable in tasks described both as formative (targeted at improving short-term and long-term learning and teaching) and summative (focused on making a judging on what has been learned or achieved) assessment in S1 and S2 data. For practitioners who take on roles as educators and assessors pedagogic concerns are high on the agenda. An AOP may be used to 'develop rapport within a class' (SLQ data) and can serve the purpose of creating not only student-centred but student-driven learning. As the student takes central position in the class they disseminate information; in doing so they take on a teaching role, even in an assessment. The range of functions and skills assessed in AOPs on taught degree programmes in the UK has not been captured in empirical literature on AOPs. These findings demonstrate that the purpose to assess was a core property of the AOPs in my study but is less distinct than at first imagined.

Taking examples of AOP tasks will expose cracks and a working definition or understanding of a genre will shift. How an individual conceptualisation fares depends on the clarity in the genre ones creates (I created AOP assessments in HE category) and what properties may hold it together. The number of properties diminish or swell as one encounters and compares different examples of AOPs with each other. What the data from S1 and S2, and this exercise in analysing an attempt to describe the AOP construct, demonstrates is that there is not a fixed or 'natural' notion of an AOP and what it involves. This accords with Adena Rosmarin's notion of genre:

> That once genre is defined as pragmatic rather than natural, as defined rather than found, and as used rather than described, then there are precisely as many genres as we need, genres whose conceptual shape is precisely determined by that need.
>
> (Rosmarin, 1985, p.25)

There may be patterns of properties in oral presentations which have been identified in previous literature on AOPs such as prolonged monologue (Hyland, 2009, Swales 2019), use of visuals (Ming, 2005), dialogic interaction after a monologue (Swales, 2019), and contribution to an academic debate and verifiable information (Levrai and Bolster, 2015). However, not all of these will be shared in all instances of AOPs. AOP tasks are complex entities which take on idiosyncratic and organic shapes, meaning any two AOPs could struggle to find points of similarity. This is an example of Wittgenstein's concept of family resemblance. There is no set of core criteria for the use of such a concept as 'oral presentation' or 'academic oral presentation'. Instead, there is 'a complicated network of similarities overlapping and crisscrossing: sometimes overall similarities, sometimes similarities of detail' (p.32). Attempts at defining and confining concepts such as these will be flawed: interpretations differ and exceptions to the given description will exist or materialize. Swales (1990) critiques this approach as 'undisciplined chaining' and irresponsible (see 2.6 for discussion). There is a negotiation of meaning and practice in particular spatio-temporal settings between non-human and human actors through speech and intra-action which does not lend itself to fixed categorisation. The AOP is in fact used and created, and in the process our notions of the AOP are forever evolving.

Leaving a construct such as the AOP as vague, without particular form is not entirely constructive – a dose of pragmatism was and is needed to make sense of AOP practices. Although defining the genre of an AOP proves problematic, it is useful to have a reference point, which can be altered and negotiated. The most steadfast of features for an instructive reference point, from the tasks I have been exposed to, are a monologue (with opportunity for preparation) communicating information within a limited timeframe to an audience, often with the use of supporting multimedia or resources. Our critical understanding of academic oral presentations can be best developed from gaining exposure to a variety of task formats, texts, their stated purposes, valued features, challenges and positives experienced with such tasks and individual applications. The genre is created for us to navigate practice but can cease to exist or be recreated in an instant.

> a text cannot belong to no genre, it cannot be without a genre. Every text participates in one or several genres, there is no genre-less text; there is always a genre and genres, yet such participation never amounts to belonging (Derrida, 1981, p.61).

In this research, I delimited the investigation to one of AOPs which acted as an assessment tools. I created the genre of the AOP assessments in HEIs in UK universities for my own research purposes but notions on what is included in this may alter so the AOPs cannot rest too comfortably. AOP is a construct on which this project rests and from exploring the concept and practices associated with it I have gained insights into the fluid and emergent concepts of genre, AOPs more broadly and AOP assessment tasks in HE.

While there is much that shifts, there is stable ground to build theories related to AOP practice (even if to be chipped away at by readers). S1 data paints a picture of tasks which develop nuanced conceptions of AOPs, and provides closer understandings on what the AOP task practices may indicate about the role of communication in HE. The use of an AOP task involves, in all cases in S1 and S2 data, the delivery of a monologue, which encompasses verbal and non-verbal communication skills. The next section considers the oracy demands placed on students and teachers in AOP tasks.

8.3   Oracy Demands in AOP Tasks in HE

This section covers how AOPs are used, thus addressing RQ1 'In what ways are AOP assessments used in UK HE?' I focus specifically on their use related to the development of oracy skills in HE but also point to skills outside the remit of oracy which AOP tasks involve.

The S1 survey findings give credence to claims that AOPs are commonly used across HEIs in the UK as formal assessment tasks (Huxham, Campbell and Westwood, 2012; Joughin 2007). The data demonstrates that AOPs are used in a variety of institutions in a wide array of disciplines (see Appendix D for disciplines represented). The survey responses from the SLQ demonstrate that AOPs are used on degree programmes in a range of disciplines from Archaeology to Art History to Medicine (see Chapter 4). It can be inferred from SLQ responses that one reason for this widespread use of the AOP seems to be the ability of the task to meet numerous objectives including the development of graduate attributes.

Graduate attributes reportedly developed at UK universities typically include aspects such as 'verbal communication and presentation' and 'interpersonal skills' (University of Edinburgh website, 2019). The graduate attributes often involve the skills needed to assemble and communicate information, not restricted to disciplinary knowledge bases. University staff are often tasked with ensuring they shape 'global citizens' who have sophisticated communication

skills and discipline-specific expertise. How can universities meet such a wealth of objectives? An AOP assessment it seems is used to begin to meet such wide-ranging objectives.

Subject lecturers cited diversifying assessment and developing graduate attributes as reasons to enlist an AOP as an assessment task on their modules. The AOPs used in an assessment context met a variety of goals, at different levels of success, according to the SLQ respondents. When the AOP is implemented in ways which align with the objectives of the module, the AOP can be a powerful genre to exploit. Teacher assessors, who have AOPs as an assessment task, note its potential to develop a range of vital life skills in and beyond the students' studies and the academy. Lack of 'constructive alignment' (Biggs, 2003), however, is an issue which may blight the AOP tasks when they are viewed as meeting graduate attributes but incompatible with the module content and objectives (see Section 4.3.1). Despite a purported lack of constructive alignment reported by 3 lecturers, the AOP is nevertheless used, potentially as it meets other criteria deemed important in UK HE. The use of an assessment to develop graduate attributes creates tension when SLs are not experts in those graduate qualities targeted in the AOP assessment, however (see Section 4.3.1). On the other hand, the AOP is depicted as a versatile assessment task, affording a wealth of educational opportunities.

The role of language and communication -- which are involved in processing and producing academic discourse, and clearly visible in university graduate attributes -- were explored in this study. Case studies which include investigation of oracy demands involved in AOP tasks in tertiary settings have been conducted (Heron 2019; Doherty et al., 2011). From consulting literature on oracy in HE (see 2.2), comprehensive data on HE practitioners' perspectives on language use and communication skills in assessed AOP tasks gathered using qualitative survey methods across institutions have not been gathered before, however. The S1 data analysis indicates that HE practitioners place differing emphasis on the means of communication in the AOP (see Section 4.3.2 on emphasis on content and delivery). The data from S1 and S2 demonstrates that fulfilling an AOP assessment task involves a range of abilities for staff and students. According to the questionnaire responses from teaching staff on subject modules, the AOP tasks may assess multiple of the following broad areas: disciplinary content skills, research skills, communication skills, interpersonal skills including teamwork, ability to assess peers and the ability to self-assess. While one SL stated explicitly that they do not assess presentation skills, other lecturers communicated the importance of delivery aspects (see Section 4.3.3). These findings provide empirical evidence of the versatile uses of AOPs which need to be borne in mind when setting AOP tasks and in EAP instruction which may claim to prepare students for AOPs in the target domain. One of the reasons cited for using an AOP on subject modules and EAP courses is that it is instantly recognizable, but the diverse tasks accessed in this study suggests otherwise.

Particular oracy skills are depicted as essential or unimportant to the delivery of AOP assessments on many subject courses by individual SLQ respondents. One respondent was keen to highlight that they 'do not assess presentation skills'. However, the responses to questions on important features in the SLQ included oracy skills such as 'audience engagement', 'clear communication of ideas', 'grammatical correctness' and 'structure' (see Section 4.4.3). These S1 findings indicate that a number of subject lecturers hold expectations regarding a degree of sophistication in oracy and communication skills. Language use in AOP assessments receives different levels of emphasis and treatment according to reports from subject teachers in the SLQ in this study. The questionnaire for subject lecturers contained a question specifically on language use in the AOP assessments (question 3.5 in Appendix A). 30 out of 45 lecturers reported they had given feedback on linguistic features, which suggests the importance of linguistic skills to particular AOP tasks (see Section 4.3.3). The wealth of complex oracy skills which an AOP may demand begs the question as to how students are supported in developing these skills ahead of the task event and whether oracy skills impact assessment decisions. Doherty et al. (2011) warns of the 'hidden assessment of oracy' in group assessment tasks in tertiary contexts including AOP tasks. The findings from S1 highlight the need for ethnographic research into the oracy demands and expectations placed on students in assessment tasks in tertiary contexts in the UK.

S1 findings not only related to oracy expectations on subject modules, but also explored oracy skills required in EAP AOPs. The EAPQ data analysis demonstrates that current EAP AOP assessment tasks may test primarily discrete linguistic skills through the use of an AOP task, while others assess a range of non-linguistic and linguistic skills (see Section 4.4.3). McNamara (1996) presents two approaches to second language performance assessment: the strong sense and the weak sense (terms glossed and discussed in Section 2.8.2). Weak second language tests consist of assessment criteria heavily linguistic in focus. There exist EAP AOPs for which less than and more than 50% of weighting is devoted to linguistic features in the criteria (see Section 4.4.3). EAP AOP assessment tasks, therefore, from analysing criteria, are currently weak and strong second language performance assessments. However, the strong and weak sense of performance assessment, while a useful concept, is based on the assumption that linguistic skills do not enter significantly into judgements into what counts in the target domain.

In comparison to subject modules, on EAP modules, language ability or communicative competence is explicitly part of the target of the assessment as well as the medium. However, while this difference in the emphasis on language is palpable, the role of language may be more pronounced in the criterion (target domain) than conceptualized in second language performance literature. This problematizes the applicability of the strong and weak performance assessment dichotomy. Language may not be included in criteria in the criterion situation but this does not

mean it is not valued by the assessors and included in score decision making. A key finding from this study is that assessors refer to aspects not included in the rating scale which is a phenomenon which has been detected in other studies (e.g. Leung, 2007; Leung and Teasdale, 1997). This is a key finding but remains somewhat marginal in assessment literature with assessment researchers continuing to validate criteria removed from the context) in which it operates. Ginkel et al. (2017) used a panel of experts to validate a rubric for assessing oral presentation performance regardless of discipline and not accounting for the social unit (Moss, 2003) in CBA. This strengthens the case for intensifying ethnographic research which investigates HE practitioner decision making, specifically treatment of language, oracy skills, and communication in the criterion situation (UK HE). This will do much to flesh out, or dismiss, any potential caveats with applying McNamara's concepts of weak and strong senses of second language performance assessment. Determining strong and weak senses of performance may also be compounded by the integrated nature of language and content depicted as 'two sides of the same coin' (Gillet, 2013). It has been argued that separating learning how to study from disciplinary content and practices is not possible and ill-advised (Wingate, 2006). And Gillet (2013) argues that language experts have little to add in a summative capacity in assessment as content expertise involves language knowledge. However, while it may be problematic separating language and communication from content, particular oracy demands may remain less visible to certain HE practitioners. Therefore, there is arguably a role for those with an expertise in language and communication for decoding and demystifying academic discourse and performances.

The findings in this research related to oracy demands in AOP assessment tasks in HE indicate that AOPs are worth investigating and understanding more closely, especially when used in high-stakes assessment. The role of oracy skills, language use (including written text), multimodal competence and communication more broadly warrants further investigation. While the notion of oracy skills used in this project (expanded from Mercer et al.'s 2017 Oracy Skills Framework: see Section 2.2) accounts for a good degree of the skills required in AOP tasks it became too restrictive a frame to investigate EAP AOP assessment decision-making practices. I used a broad notion of oracy in this study (see Section 2.2), however what counted in EAP AOP assessment moved beyond the realm of oracy skills. This is now discussed in relation to EAP practice and in doing so Research Question 2: 'What is the EAP AOP Assessment Construct?' is addressed.

8.4   Complexity of EAP Construct in AOP Assessment

The parameters of the AOP construct were not outlined prior to setting RQ2. In asking RQ2, I attempted to clarify what the practices indicate about the construct in AOP educational assessments. The study did not seek to establish construct-relevant and construct-irrelevant

variance. In fact, these notions do not map neatly onto the processes in FS1 and FS2 when there is scope for emergent organic practices and teachers and students have an expanded relationship. It is highly problematic for a researcher to enter such a setting and evaluate the assessment outcomes; I sought instead to describe and theorise what the AOP involved and discuss the potential ramifications of this for notions of EAP practice, specifically related to AOP tasks.

The AOP assessments on EAP courses are primarily labelled as 'speaking assessment' by the EAP practitioners at the two S2 field sites and at EAP practitioner events (such as the Speaking PIM held in 2018 by the EAP professional body BALEAP). S1 data analysis demonstrated that some EAP AOP criteria continues to focus purely on spoken linguistic features, however the majority of EAPQ respondents and presenters at BALEAP's PIM communicated that they were implementing integrated AOP assessments, which encompassed a range of skills including an array of academic literacy and oracy skills. As discussed in Section 8.3, S1 data on what criteria (in the rating scale) is used to assess AOPs indicates that, in particular contexts, AOPs may assess primarily linguistic skills associated with speaking, while in other AOP tests the criteria are a more equally weighted integration of a range of communication skills and academic skills (see Section 4.4.3). Study two builds a more complex picture of practices, however, as the criteria in the rubric is supplemented with, or supplanted by, what teacher assessors value (see Sections 6.3-6.8). The teachers in their rater conferences and interviews reveal a wealth and range of qualities, which are regarded (to various degrees) to be important components in the assessment tasks.

The ethnographic fieldwork in S2 of this project delved deeper into practices and found that the ways AOP assessments are described and conceptualised does not reflect the complex and sophisticated constructs EAP practitioners may be using. While the AOP was labelled a 'speaking assessment', the teacher assessors had a multifaceted notion of what inferences were being made about the student test takers in the AOP assessment process. The core construct the teacher assessors were working with in these gateway AOPs was the students' ability to cope in their prospective academic studies in the medium of English, which involved a vast array of target abilities and concerns.

The S2 findings show that the EAP practitioners use the AOP assessment as an opportunity to gather evidence on a range of skills which provide indications as to the student's ability to cope in the target domain. Fulcher and Davidson (2007) describe the ability to cope as a subject lecturer evaluation of whether students can perform tasks. Previous research has neglected to investigate whether, and how, the ability to cope is used as a construct and what this encompasses in AOPs. Research on how 'subject lecturer evaluations' would be determined and neatly packaged in a construct for an AOP task need to be carefully considered. Would AOP tasks in the target domain

and expectations be used as a basis for decisions in AOPs or broader qualities required in HE? How do identities and issues at the level of epistemology shape these practices and how would this be reflected in an assessment construct (discussed also in Section 8.6.3)? The teachers may use the term ability to cope because it describes how they look for the minimal skills needed to function on their prospective courses. It is used so as to ensure they do not deprive students of an opportunity to *enter* Higher Education unless absolutely necessary when reaching gatekeeping decisions.

Teacher assessors look for a complex array of knowledge and skills in an AOP performance on EAP programmes. Januin and Stephen (2015) identified 'public speaking, use of the oral presentation structure, and linguistic knowledge' as part of the discourse competence required of students on an EAP AOP task set at a university in Malaysia. The Survey Study and The Fieldwork Study findings, in this project, demonstrate that the abilities required in delivering a successful AOP from teacher assessor and student perspectives are far broader and complex in the settings accessed in this project than found in Januin and Stephen's study. This could be explained by the fact that the task was not described as an assessment task and had a specific learning focus to develop a plan for an essay in Januin and Stephen's context. The teacher assessors in S2 comment upon cognitive skills, communication skills, social and emotional skills and formative learning skills. A complex assortment of skills is evaluated from evidence within the AOP, but the assessors do not limit themselves to the confines of the AOP performance when making evaluations and reaching score decisions.

It is not only what is contained in the AOP performance that provides evidence of student abilities to the teacher assessors when reaching score decisions. The AOP acts as a springboard to consider information garnered from assessment opportunities outside of the AOP formal assessment event during the lifespan of the course (see Sections 5.6.1, 5.6.2, and 6.7.2). The teacher assessors draw on evaluations of students made prior to, during and after the assessment. The judgements reached during rating discussions relate to the person, and are not restricted to the performance (see Section 6.7.2). The performance will not be entering the academic realm, but the person performing. Torrance and Pryor (1998) also found an inclination among teachers to use knowledge of the 'whole' student when interpreting assessment opportunities, when assessing young children in classroom assessment. The teacher assessors in S2 draw on the assessment opportunities they have created and exploited over the course (for reasons of validity and fairness– see Section 8.6 below). The construct is complex in the educational settings in which the assessor and test takers have had (in this case prolonged) prior contact. The complex decision-making processes have been researched in EAL in primary education (Torrance and Pryor 1998; Rea-Dickins, 2001) and benefit contemporary EAP pedagogy and research.

Context is part of the construct, making a construct in educational assessment difficult to predetermine and control. Fulcher and Davidson (2007, p.5) argue that a construct such as ability to cope needs to be measured and defined and this is a core validity concern for tests which assess readiness for academic study. However, on educational programmes which involve close and prolonged contact with small groups of students, the ability to cope in a tertiary education environment will be gauged in numerous assessment opportunities open to the assessors. This proves difficult to anticipate or orchestrate. It proves problematic to map an AOP performance to a fixed rating scale to make the most valid and fair judgements possible (see Section 8.6 below on validity).

The construct is further complicated in the AOP educational assessments as the assessment process involves the ability of teacher assessors and student test takers to co-construct the task representation in an evolving, dynamic setting (Pryor and Torrance, 2000). In S2, the teachers expect students to be susceptible to continual cues emanating from them before and during the assessment event on what is required in the AOP tasks (see Sections 5.5 and 6.7.3.4). This demonstrates that part of the construct is a sociocultural phenomenon of co-constructing task representations. This finding echoes Wolfersberger's (2013) findings that the student task representations are important to successful performance in classroom-based assessment on reading-to-writing tasks. Moreover, it gives credence to Wolfersberger's (2013) stance that the students' conceptualisation of the task should be a feature of the assessment construct. The complexity is heightened further by the findings that teacher assessors and student test takers may have different task representations (see Sections 5.4 and 7.2.2 in particular). This coheres with Salter-Dvorak's (2016b) findings that students and teachers may have conflicting values in AOP tasks. Teachers, in the rating discussions, have the opportunity to negotiate such ideas about the task and the construct in the rating conferences. The students should be afforded a similar opportunity to express what their task representation is and the rationale behind their approach to the summative assessment (discussed further in Section 9.2.3 and 9.4) .

From exploring what the construct is (thus addressing RQ2), the findings have contributed to our understanding of what EAP assessments can involve and developed understandings of the construct of EAP. These insights, based on empirical data, address the gap in research in EAP assessment which contributes to conceptualizing the 'under-defined' construct of EAP (Hamp-Lyons and Schmitt, 2015). The survey and ethnographically-oriented fieldwork have developed an understanding of what the AOP educational assessments involve. The findings strengthen the argument in CBA research that context is part of the construct (Moss, 2003). The fluidity of the teacher and student diverse task representations complicates the notion posited by Fulcher and Davidson (2007) that a construct such as ability to cope can be measured and defined, in

educational assessments. This opens up monumental questions as to how high-stakes assessment should be done in educational settings which involve multiple and co-constructed assessment opportunities. The teacher assessors show there is a delicate balancing act between developing learning and retaining a focus on the gatekeeping purposes of the summative assessment tasks.

One way forward is to aim for flexible reference points sensitive to a range of evidence available in contexts and enhance teacher assessor training (see Section 8.6, on validity and fairness, for further discussion). Crucial questions persist as to what tensions or pitfalls may surface if the summative dimension is diluted to a large degree and the assessment opportunities broadened (see more discussion in 8.6).

Approaches to EAP assessment were shaped by the orientations towards what role language plays in the ability to cope in academia. What role it was assigned in the AOP task also significantly factored into assessment decision making.

8.5   Role of Language in EAP AOPs

The students on EAP courses described a process of transitioning from courses and tests which focus on English language proficiency to their EAP tests which require a range of abilities which move beyond linguistic competence. English language tests in secondary education in students' home countries, or standardised tests such IELTS, are described by participants in this study as showcasing mainly English proficiency. By contrast, EAP students are often required to display a complex set of skills required to perform an academic task in EAP integrated performance tests. The students, therefore, dedicate a large part of their energies to developing and showcasing academic skills (see Section 7.3.1). Due to this shift in focus, in certain cases, students notice a marked change in priorities: the students regard linguistic performance as one part of their EAP AOP performances. In relation to assessment of language, the students take up the AOP as an opportunity to develop language, but not to particularly showcase it (see Sections 7.4 and 7.8). This suggests that the formative dimension is stronger than the summative dimension in relation to linguistic performance. Adopting McNamara's (1996) terms, the students conceptualise IELTS as a weak second language performance assessment and the EAP AOPs at the two field sites are, in comparison, a strong second language performance assessment. Students' descriptions of the assessment construct depart from what is indicated in the rating scales. This may be due in part to the teacher assessors' communication in class and feedback and also the students' perceptive observations as to what warrants the most attention to fulfil the AOP task requirements.

The students move from a linguistic focus to a consideration of what the construct of successful communication looks, sounds and feels like in their AOP tasks. The students cited using a range of

semiotic resources such as gestures, visuals, body language and the modality of PowerPoint slides (see Section 7.5.2). Previous research has found that second language users dedicate a substantial amount of time to preparing slides in order to cover current 'gaps in their linguistic repertoires' (Yates and Orlikowski, 2007). This coheres in part to the findings in this research. The students stated that if they did not have particular lexis or syntactic structures as part of their personal repertoires, they could use other semiotic resources to compensate. A number of students at both sites also, however, argued against viewing the verbal communication as necessarily central in the communicative act (see Section 7.4.1). Therefore one must exercise caution in attributing certain use of slides as always covering 'gaps' in linguistic repertoires. Speech is not treated in isolation from other communicative resources by the students; they considered the whole embodied activity of the AOP. These students reflect the poststructuralist view of communication, conceptualising communication as an assemblage (Canagarajah, 2018: see discussion in 2.8.4). Assemblage describes a process when 'all modalities including language work together and shape each other in communication' (Canagarajah, 2018, p.39). Whether other stakeholders take this view, has major ramifications for the student test takers (see implications in Section 9.4). The emphasis on slides and verbal communication needs to be negotiated and clearly communicated in AOP assessment tasks.

The teachers take a similar perspective to the students, but place stronger emphasis on the linguistic skills. These experienced EAP practitioners have a conception of a base level of linguistic performance that they feel is required to function successfully in the target domain. The teachers acknowledge that a number of the students on their courses already have a sufficient base level of linguistic skills, for example Marina in FS2. The teacher assessors discuss linguistic performance in the rating discussions, however this is often eclipsed by a focus on communicative performance. Research on the role of language as predictor of academic success reports that it is probable that once a particular level of proficiency has been attained, then other factors become more central to successful academic performance (Graham, 1987). What this 'base level' is, however, is difficult to pin down. Each interaction entails a new strategic blending of communicative resources. The teacher assessors view language through a communicative frame on many occasions, but they isolate language from the other semiotic resources in line with the rating scales mapped to AOP performances. There is a shift away from a linguistic orientation, but not to the same extent as expressed by a number of the students.

A factor shaping the shift away from linguistic orientation in practice for teachers is that many practitioners approach the treatment of language from a communicative frame (as shown in Section 4.4.4) and experience difficulty in noticing language use when focusing on the integrated AOP communicative act. The teachers in S2 and respondents in the EAPQ communicated their

difficulty in noticing language use, especially grammar and accuracy (see Sections 4.4.5 and 6.3) in their AOP assessment tasks. Statements on the challenging nature of noticing accurate and complex spoken language use suggests that this is difficult to catch unless it is the primary focus of the raters' attention. The zooming in and out strategy of checking close-up interpretations with holistic ones (see Section 6.3) was crucial to appraising the validity and fairness of evaluations of language use. The teacher assessors in FS1 also expressed discomfort and a dislike of having a separate language section in the rating scale (see 6.5.1). There is a fundamental concern as to whether the language descriptors match the communicative task at hand. The AOP events involve a range of semiotic resources, but the AOP assessment criteria and university policy requirements are constraints (see Section 5.3.3) which may nudge the teacher assessors to extract and evaluate language use. The rating scales at the two sites encourage teacher assessors to evaluate a range of skills, but to also keep a focus on the linguistic. EAP practitioners are understandably aiming to mentor students in a range of complex skills to help them navigate the academic realm. However, in the case of the AOP tasks, the very act of covering multiple abilities and focusing on communicative performance makes it near impossible to evaluate linguistic skills in isolation. A question arises as to whether teachers should force noticing language use? Would it be noticed in the target domain potentially by forgiving or unforgiving interlocutors? Do teachers notice linguistic features which are communicatively salient?

Widdowson (2020) argues that English learner perspectives on what students notice and regard as salient to communication should inform language learning curricula. He also seems to present teachers as inculcating native speaker norms. In this study, the perspectives from students provide valid insights on the realities of language use and communication in AOP tasks and offer insights into what is valued in AOP communication. Teacher perspectives are not necessarily as removed from communicative reality and out of line with ELF perspectives as Widdowson portrays. In current EAP assessments, teacher assessors are engaging with questions as to what warrants admittance into assessment decision making and diluting the role of linguistic features in certain circumstances. Language is not unimportant and teacher assessors navigate complex reasoning behind focusing or not focusing on linguistic features of performance.

The data in the study contributes to an ensuing debate on focus on form in language education and testing (see Section 2.9.2). Students noted that an attention to form in the written language on the slides was important (see Section 7.5.1). Daniel stated that focus on form is necessary as divergences can delay processing, especially in a real-time AOP performance when successful communication is contingent on the ability to digest the information communicated with relative ease. A focus on form cannot be completely eradicated -- as Chopin (2014) argues -- in AOP events, but ways in which resources are assembled which may render particular instances of form

unimportant must be acknowledged and rewarded in performance assessments which test communicative performance. In the EAP AOPs, there is a gradual move away from linguistic focus to a communication focus in the implementation, in line with practices advocated by translanguaging, ELF(A) and critical language testing research (see Section 2.9.2). While teachers focus on form they use a zooming strategy (see Section 6.3) to establish how any inferences based on discrete language use match with judgements made at a holistic level on the students' communicative success in the AOP performance. In these assessments, the assessors reject the notion of native speaker competence as the benchmark. McNamara (2014) appeals for a radical reconceptualization of the construct of successful communication. The data in this study indicates that a construct of successful communication which moves beyond a focus on the linguistic gravitates us away from the native and non-native competence distinction. The power of an effective assembler of communicative resources can be rewarded in AOP assessments which encourage the use of diverse modes and semiotic resources. These findings as to how an AOP performance is approached and evaluated give us fruitful insights into how communication works in similar and different communicative acts to the AOP and directly responds to McNamara's calls for work on reconceptualizations of successful communication.

The explicit recognition of assemblage as being part of process and fundamental to the construct is somewhat lacking in EAP AOP assessment. In an assessment situation, it would be detrimental to the students' scores if teacher assessors marked based on a structuralist approach rather than an integrated, holistic approach to communicative activity. There is a movement away from the structuralist approach (see Section 2.8.4 above) but it is still present. The notions of assemblage and spatial approach to communication have great explanatory power in this research. If language is required to be isolated, then an AOP task which can afford minimal use of language would need to be seriously reconsidered. At the very least, I would suggest that the use of visuals would need to be rethought. Alternatively, the AOP can be embraced as an activity which allows students to showcase their abilities as assemblers of communicative resources.

The discussion in this section highlights the need to potentially move away from labelling and conceptualising EAP AOPs as speaking assessments if there are other modes of communication available. Even if verbal monologue is a 'privileged property' of AOPs this does not make it central to specific tasks and to a particular presenter and audience member. The AOP practices in the two field sites show that students and teachers use AOP assessments to display, develop and evaluate communicative performance. The assessments and the gatekeeping decisions are also shaped by an assessment of content-related skills indicating a level of capabilities needed to produce academic discourse (see Section 6.7.3. for qualities of 'ability to cope'). The teachers consider how ready the students are for challenges that await them in the domain of academia beyond general

academic English language proficiency. The teacher assessors look for signs of communicative capability, academic aptitude and how students function on an educational course. There are multiple layers of abilities tapped into and it proves difficult to encapsulate this complexity into a neatly packed construct and label for the assessment, especially as individual EAP units and practitioners take different approaches. The AOP assessment becomes blended with other formal assessments and informal assessment opportunities during the course's lifespan. The target and construct of the assessment suite as a whole, as well as individual tasks needs to be borne in mind. A more fitting label for the EAP AOP assessment could be 'academic oracy performance assessments'. Although what is valued and assessed is not restricted to oracy skills (see Sections 8.3 and 6.7.3), this is a broader frame than 'speaking' and can incorporate the bulk of shared EAP AOP practice featured in this study. No label encompasses all facets of practice under its umbrella.

Spoken communication is one resource drawn upon and spoken utterances must be evaluated in conjunction with other semiotic resources used. Otherwise, there is a detrimental effect on those taking advantage of the assemblage of resources the AOP tasks not only afford but encourage. A major validity concern has been raised as to the how feasible and fair it is to isolate linguistic skills in particular AOP tasks. Validity concerns have been touched upon during the thesis; I now turn to an extended discussion of this pivotal topic in educational assessment.

8.6   Validity and Fairness in AOP Performance Assessment

Despite a wealth of CBA practice, including the use of AOP assessments in EAP settings, there is a paucity of research which describes the decision making in teacher assessments in a way which contributes to debates on the appropriate methods for validation in such contexts (Bonner, 2013). There have been few studies which move beyond self-reporting techniques and access and analyse decision making between practitioners during AOP assessments. This study adopts an interpretivist stance to the investigation of assessment practice. It deepens understanding of gatekeeping decision making through gathering data from stakeholders who are well-versed in the workings of their contexts. The decisions teachers make are qualitative and iterative in CBA (Kane, 2006), but there is still much to understand in terms of the techniques and strategies used by teachers and students in decision making (Bonner, 2013), especially in the marginalised genre of AOP tasks. There is a risk that actions may be misjudged as justified or unjustified if such critical inquiry is not undertaken.

 Researchers differ on their stance as to the place of traditional psychometric validation techniques, applied in standardized testing, in CBA. Bonner (2013) argues that traditional

approaches can contribute to enhancing validity (excluding decontextualisation, standardization and pseudo-objectivity), whereas Brookhart (2003) believes completely new theories must be applied to CBA. Pamela Moss and Susan Brookhart advocate that assessments which involve close and prolonged contact between teacher assessors and student test takers on educational courses are most compatible with validity approaches which are underpinned by social-constructivist epistemology. The role of context and co-constructed nature of assessment practice are key drivers in learning and assessment decisions taken by students and teachers. Teacher assessor descriptions of the assessments in the field sites, initially, were ones which indicated that the AOP performances were mapped onto a rating scale and assigned a score accordingly. The practice reveals a far more complex phenomenon takes place which is less standardized than the teacher assessors themselves envisage. The complexity derives from characteristics of CBA which are shaped by the social dimension of practice. The validity of the classroom-assessment featured in this project is discussed in reference to the following factors: dual learning and measurement assessment purposes (Section 8.6.1), prolonged engagement with students (Section 8.6.2), multiple sources of evidence available (Section 8.6.3), importance of considering multiple perspectives (Section 8.6.4), strategies to be tapped into as human assessment instruments (Section 8.6.5), the need to seek reaction from co-assessors (Section 8.6.6), the argument that subjectivity can be a virtue in assessment (Section 8.6.7) and the acknowledgement of official and unofficial practice at work (Section 8.6.8).

8.6.1 The Learning and Measurement Purposes

The learning as well as measurement purpose of the educational instruction and assessments encountered in the tasks which feature in this study complicate assessment decision making. The AOP tasks involve measurement in assigning a score used for gatekeeping purposes and simultaneously are used to further learning. The assessments become an intervention tool and track development, meaning the assessment actively seeks to be obtrusive, unlike in traditional assessment (Bonner, 2013, p.98). I call the informants in this study teacher assessors and student test takers because the educational purpose is the lifeblood of the vast majority of actions including the substantive processes involved in summative teacher and student assessment decision making. The formative (assessing for learning) orientation of practice shaped student and teacher behaviour to a large degree (see Sections 6.7.3.4, 7.8) and the assessment event itself was a critical learning episode for teachers and students (assessment as learning). The learning purpose is shaped by the context and actors: a central tenant in teachers' pedagogic values is their desire to cater to the individual needs of individual students. This inevitably involves flexibility and modifications to practice, many of which may be unforeseen.

8.6.2   Prolonged and Close Engagement

It is not only the learning purpose, which may justify much of the evidence and strategies summoned by teacher assessors when evaluating students. It is the time, changes to context and actors, and the nature of contact which eclipses other considerations. Contextual effects are not strongly controlled in CBA and it is acknowledged that the interpretive arguments made can be altered over time with new and compelling evidence (Kane, 2006). Relevant differentiation in instruction and evolving expectations occurs for teachers, from student to student, year to year and over the life span of a course over an academic year. Therefore, such differentiation is surely to be reflected in assessment episodes. Bonner (2013) poses the question: 'If instruction is differentiated, shouldn't assessment be differentiated, accordingly?' (p.91). Validity arguments must take account of this differentiation, a characteristic avoided in large-scale assessment through standardization.

8.6.3   Multiple Sources of Evidence

The teacher assessors in FS1 and FS2 show signs of differentiation in the assessment process. The prolonged and close contact which affords a wealth of assessment opportunities and insights into each individual students' capacities is the powerful tool which gives weight and plausibility to claims made as to the score the student receives. If deemed unrepresentative of a students' abilities, one performance is considered a particularly poor piece of evidence and performances outside this official formal assessment episode are granted entry into the assessment decision making proceedings by teacher assessors. The unit of measurement becomes the social context and the student instead of one performance. Bonner (2013, p.96) notes that tests which involve an element of predicting future performance often draw on information from outside sources and 'external variables' such as effort. Effort is presented by teacher assessors as a sign of impetus to cope and therefore a core quality, not an extraneous variable.

In this study, teacher assessors often gathered evidence from unofficial sources, from informal methods such as observing classroom behaviour, and other formal assessments. The triangulation of different windows on student performance is a strategy which is used in the rating discussions. For instance, the teacher assessors consider how the student will fare in probable scenarios in the target domain which are not sampled in the AOP assessments (see Section 6.7.1). This is used as a means to better ensure that the test decisions reflect the ultimate gatekeeping decision that is being made. In this study, external information seems to be used to make sure a poor AOP performance score is justified or to increase a score, not to lower a score. The teacher assessors regard this decision to be whether *the student* has what it takes to cope with demands in the

target domain. Assessing the person and not purely the AOP performance is important to the EAP practitioners (see Section 6.7.2) in ensuring the inferences made as to how students will perform in the target domain are as valid and fair as possible. Tracey (FS2) says, in member checks, that she has 'wiggle room' to bring in knowledge of a student external to the AOP performance when it increases the validity of inferences made as to who will fare well on their prospective courses. Teachers bring in knowledge of social factors (see Section 6.7) gathered from interactions with the students. Broadening the scope of the interactions and context admitted into assessment is valued as a means by which to make valid inferences as to who will cope in the target domain. The multiple sources are also used to consider individual approaches to tasks. Wolf et al. (1991) advocate being open to 'multiple paths to excellence' (p.63) and not a 'single progression from novice to expert' depicted in rating scales.

The decision to hand the student the key to the gates of UK academia ultimately rests on a patchwork of judgements and how these are arranged into as holistic a portrait of the student as possible. The teachers consider the effects of the decisions they make by theorising how the student would cope and the effect on peers and disciplinary tutors in the target domain. To establish the validity of such practice, research is needed which investigates whether the qualities teacher assessors look for in fact lead to success in academic study. Furthermore, the student test takers should be made aware that performance throughout the life span of the course could be drawn upon in assessment. This may raise issues of performance fatigue which would need examining further. If all interactions are explicitly labelled as 'assessment opportunities' would this negatively impact learning and rapport between teachers and students? Ongoing assessment could be interpreted as relentless surveillance and create a tense and unproductive learning environment with no space for trial and error. An open dialogue between teachers and students would be required to establish how such harmful impacts on learning, fairness and validity would be minimized.

8.6.4   Multiple Perspectives

The prolonged interaction between student test takers and teacher assessors entailed a complex co-construction of the assessment task. The qualities of the social setting and actors become crucial to determining validity and therefore students' thought and action related to assessment episodes is of paramount concern. Student and teacher interpretations of the assessment and assessed performances diverge at times (see Sections 6.8.1, 7.2 and 7.6 for examples). The students encounter the task first hand and arguably have a closer understanding of what it requires to deliver. This signifies that students are entitled to be involved in determining what is deemed valid. This could involve reflective components to the task which afford students

opportunities to justify their approaches. Negotiations and debriefings with students can illuminate potential areas which could improve instruction or increase fairness in rating practices. For instance, the interviews with students in this study have highlighted validity concerns as to how language is treated in rubrics and by teacher assessors in AOPs on EAP courses. Student therefore should, where possible, be included in validation processes.

8.6.5   Maximising Human Instrument

This study contributes to our understanding of the range of techniques and strategies available to human raters which could be further investigated in future research and harnessed by practitioners to enhance validity and fairness. The assessors employ a range of lenses (descriptive, evaluative, judging, marking, technical, aesthetic), a zoom function and put the formal AOP performance through a variety of filters (criteria, ranking, social, interpersonal; see Chapter 6 above). In S2, the teacher assessors shared notions of what could count as features of students' readiness to commence their disciplinary studies which have been presented under the construct of 'ability to cope'. From analysing rater talk (see Chapter 6), it emerged that teacher assessors are using a complex blend of traditional and interpretive strategies. They use the criteria and the performance which retain the measurement purpose while also acknowledging the differentiation which exists in their practice by using the social filter, for instance. The teacher assessors take a sophisticated pick 'n' mix approach, endeavouring to use the strengths of diverse strategies at opportune times (see Chapter 6). The teachers lean towards strategies used in naturalistic research when enhancing validity. Johnston (1989, cited in Moss 1992) strongly advocates that teachers, when justifying judgements made on student performance, use such strategies as those in naturalistic research outlined by Lincoln and Guba (1985): prolonged engagement, different sources of evidence, seeking reaction from colleagues and leaving an audit trail. I have discussed prolonged engagement and different sources of evidence previously in this section. Seeking reaction from colleagues in the designated score-reaching discussions proved vital to investigating thinking and actions for the teacher assessors.

8.6.6   Seeking Reaction

The rating discussions between the two assessors were used as a time to seek reaction from co-assessors in order to gauge the validity of interpretations. The raters at points demonstrate assessor reflexivity when debating whether aligning with the criteria strengthens or weakens validity and whether mitigating factors are admitted to the decision making (see Sections 5.7 and 6.7.1 in particular). The teacher assessors use rating discussions to declare the justifications and sources behind their decisions. The teachers also have an opportunity, through dialogue, to build

a degree of consensus on the assessment construct. Williams (2001; cited in Leung 2005) describes construct-referenced assessment as potentially defined as 'the consensus of the teachers making assessments' and it depends on 'the existence of a construct (what it means to be competent in a particular domain) being shared by a community of practitioners' (pp.172-173). Each assessment is 'a head-on encounter with a culture's model of prowess' (Wolf. 1993, p.213). The teachers have differences in task representations, however the process reveals some common values (see Section 6.7.3). The interpersonal dimension (see Section 6.8) may have a negative bearing on the final scores if for instance one person dominates the discussion for invalid reasons (more research is needed to explore this area), but through such dialogue the assessment decision-making becomes visible and open to critique and transformation. Dedicated time for multiple raters to discuss performance and the wider context and actors is a core validity-enhancing strategy in AOP assessments in educational settings – for the current assessment and for future design and implementation decisions. The act of critical inquiry and justifying a decision is particularly important in an assessment which is less standardized.

8.6.7   Subjectivity as a Virtue

Reducing bias and increasing objectivity is advocated through use of rubrics by researchers and teachers who feature in this project. However, findings in this study and other research shows that teachers make decisions based on qualities external to the rubric, such as effort and behaviour (North, 2000; Llosa, 2008). Illuminating the justifications for including such qualities is key to establishing the validity and fairness of such practice in an environment which has not been controlled to a large degree. The teachers acknowledge that the expanded relationship with the teachers and students in the setting renders an emphasis on objectivity and depersonalisation of students' performances as unrealistic, and unfair. Georgina powerfully describes how she cannot erase previous experiences (in her reference to neuralyzers in the film Men in Black in Section 6.7.2). The best way to ensure fair decisions are made is if the subjective nature of the decisions are declared. Husserl (1977) believed that human lived experience cannot be measured objectively. The overpowering default lens which is a constant in the assessment process (and all human lived experience and action) is a lens of subjectivity. Subjectivity has its flaws, which the teacher assessors acknowledge, but to ignore the subjective experience, as Johnson (1989) argues, is to '[hide] the political nature of the activity in a veil of science' (p. 512). This becomes increasingly relevant the less standardized and controlled the interaction between teacher assessor and student test taker. The windows on capacity are fewer and the interpretive element is arguably minimised in standardized tests. The interpretive dimension is intensified with greater windows available on the capacities of students in such assessments and warrants validity

strategies to improve fairness, which recognise the subjective nature of the endeavour. The assessors must exercise reflexivity and ascertain what assumptions may needed to be bracketed out and which may be a force for good. Further close analysis on this decision making in action is welcomed. The influence of the social and the individual in the process, the human dimension of practice, needs further examination. The human instrument, when reflexive, proves it has the potential to be more flexible and humane than the rubric.

8.6.8   Official and Unofficial Practice

The findings demonstrate that alternative strategies are being implemented to strengthen the validity of EAP AOP performance assessments. However, there is a weak trace of these strategies because of a sense of guilt experienced when moving away from the rating scale and personalizing the assessment process. This is, to an extent, understandable as such assessment decision making has not been made explicit to stakeholders, most importantly the student test takers. The covert behaviour is not necessarily fully legitimized but it can be understood and recognised as useful.

Standardization of assessment can have an adverse demoralising effect, and can disenfranchise the teacher assessors in educational assessment (Van Lier, 2004). This is particularly evident when the teachers consider the ethicality of the assessment process. The guilt the assessors express when they bring in information outside of the AOP into the discussion to ensure fairness, but the conviction they express as to how it is used out of fairness seem contradictory on the surface. This is due to the assessors applying standards associated with large-scale assessment to educational assessment. The teacher assessors are careful to include 'observable evidence' and base decisions on what is 'visible' in the AOP performance and to adhere to the rating scale. Simultaneously, they also consider how representative the criteria in the rubric is to essential qualities required to function on a tertiary educational course. And they consider whether the student performance is representative of what the student presenters can do which will stand them in good stead on their academic courses. The teacher assessors may consider aspects which are difficult to trace in the rating scale because the pre-defined criteria are not considered as relevant to the ultimate decisions or the nature of assessment opportunities that have unfolded on the course (see Section 6.5). The teachers wish to assess the abilities that the course has developed and those which truly count in being able to cope in the target domain. The teacher qua assessors navigate fairness and justice implications and adopt complex strategies in the face of ethical quandaries. Their actions intensify the credibility of inferences made as to how the students will fare in their future studies. There are the official and unofficial spheres of assessment practices; the unofficial can do much to strengthen validity. Bringing the unofficial into the official domain may be a

power for good but there may be harmful effects on learning (as discussed in Section 8.6.3 above).

If such blended approaches (traditional and alternative strategies) in appropriate settings or complete naturalistic approaches (if possible in the context) were advocated, then the potential to bolster validity may increase. The detection of the complex mix of strategies, deployed judiciously, has been possible using an ethnographically-oriented approach. Moss (1992) calls for those involved in educational assessment to 'expand our repertoires of epistemological strategies and consider alternative models for warranting validity conclusion' (p.253). This research supports an interpretive approach to assessment, in which subjectivity is acknowledged. Alternatively, in situations when prolonged contact is not possible a blended or hybrid approach may be a more effective and pragmatic approach (see Section 9.4 on further research). The findings add to a knowledge base concerning what can constitute alternative models of validity in language+ assessment. It makes the case that there is a fruitful epistemological strategy in such grounded, interpretive research.

8.7   Chapter Summary

This chapter, in discussing core findings in this project, has identified the contributions this thesis makes to a range of areas in Applied Linguistics and Educational Assessment. The discussion has offered a rich description of AOP practices which provide a nuanced understanding of the diverse applications of assessment tasks on taught programmes, which are conceptualised as part of the AOP genre in UK HE. The discussion and associated data (in particular in Section 8.2) contributes to an AOP genre knowledge base of HE stakeholders. In section 8.3, I discussed findings related to the oracy expectations HE practitioners reported to have in AOP assessment tasks, which gives an indication of the oracy demands placed on students and HE practitioners. The discussion contributes to a growing body of work into oracy and communication skills required in HE.

In Sections 8.4-8.7, this chapter focused on AOP assessment practices on EAP courses. In Section 8.4, I discussed how a wish to draw on a wide evidence base of a range of abilities and large scope for differing task representations results in a complex AOP construct. The context in educational assessment becomes part of the construct, whether fully acknowledged or not by teachers and assessment developers. There is a focus on academic skills and students' abilities to function in an academic environment for which their EAP module becomes a microcosm. In Section 8.5, the conception of a successful communicator was revealed to be an effective assembler of resources – verbal communication being one of many resources. This construct of successful communication, and the difficulty in noticing aspects of language use reported by EAP

practitioners, points to a need to reconsider how language or communication is assessed in an AOP task. Section 8.6 was dedicated to a discussion on how alternative and interpretive approaches to validity have traction in EAP AOP settings, despite these being used unofficially. Through the inclusion of social factors, the subjective nature of the process is declared and has the potential to be a force for good. Validity is relevant to the local purposes of the assessment and relative.

This chapter makes the case that this research contributes to a gap in EAP assessment theory building (Schmitt and Hamp-Lyons, 2015). It also addresses lacunae in the literature concerning close descriptions of EAP assessment in practice and assessor and test taker decision making (Hill and McNamara, 2011). This project enhances debates on validity in language+ testing and educational assessment. There remains a great deal to be more closely understood, and this task is the subject of much of the concluding chapter's contents.

Chapter 9　　**Conclusion**

9.1　　**Introduction**

The previous chapter provided a discussion of findings. In this final chapter of the thesis, I set out key insights offered into AOP practice in this research output. I begin by summarising the key findings (in Section 9.2). I then outline limitations of the project (in Section 9.3) and implications for practice and further research (in Section 9.4). I close this chapter, and the thesis, with reflections on the PhD process in the concluding remarks section.

9.2　　**Key Findings**

The first key finding in this project is that the language construct in AOP assessments may not adequately cohere with the communicative affordances and characteristics of the AOP tasks. The second key finding is that teachers use a sophisticated array of tools and strategies in enacting AOP assessment which occupy official and unofficial spheres in assessment decision making. The third key finding is that teacher assessors require dialogue and resources to engage in critical appraisal of their assessment thinking and action.

9.2.1　　**Discord between Linguistic Focus and AOP Communication**

How communication may be achieved in AOP tasks requires sustained attention and to be made explicit to teachers and students. Otherwise, there is a threat to the validity of tests, particulary if teachers assess abilities which do not match the communicative reality. Students who enact the task may grasp the communicative reality better than the assessment developers and teachers. The voices of the test takers must be heard and a platform provided for them to put forward justifications behind their approaches to a task to ensure that fair assessment decisions are being made (see discussion below in Section 9.4). The cognitive and social demands of the EAP AOP tasks need to be acknowledged. The findings in this study revealed that teachers are employing lenses, filters, and zoom functions to best manage the assessment process. Teacher assessors are to be commended for handling a complex array of tools and strategies. However, there are pitfalls which are encountered in the rating process.

One aspect of particular concern, with validity ramifications, is how the assessors are able to focus on discrete linguistic features and communicative effectiveness in real-time AOP performances. The teachers switch between the t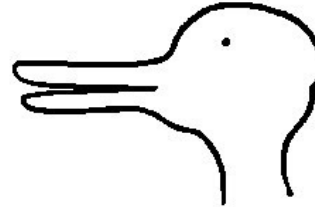wo perspectives (zoom in to specific features, and zoom out to get a holistic interpretation). This is proving challenging because, in order to capture the close-up and holistic evidence, a constant shifting of perspectives on the AOP delivery

Figure 18: Duck Figure (Wittgenstein, 1958)

Figure 19: Rabbit Figure (Wittgenstein, 1958)

needs to occur. Taking Wittgenstein's (1958, p. 194) 'rabbit-duck' ambiguous figure (derived from Jastrow) as an example the teacher assessors are switching from the duck to the rabbit to obtain evidence; they report difficulty noticing both at the same time in real-time AOPs. (See Figures 18 and 19 which have been slanted in ways to make the rabbit or duck more discernible). Switching from attention to specific features such as pronunciation or grammar to the overall communicative effectiveness of discourse was problematic. The teacher assessors extracted examples of language use which at times they only notice when honing in on language use and these ran the risk of not representing the students' overall communicative performance.

The AOP task can be used to push the bounds and parameters of language assessment due to the communicative affordances it involves. The teacher assessors' practice is grappling with the opportunities and pitfalls associated with rating AOPs, but it has been undertheorized. While English language proficiency test such as IELTS are language-oriented, the EAP AOPs have the ability to tap into more authentic communicative capability due to their educational focus and embedding in an educational setting, which is to a degree successfully exploited with prolonged and close engagement between teacher assessors and student test takers at the two sites. EAP assessment developers and practitioners have core decisions to make: use an AOP performance assessment with opportunities to use multimedia and non-verbal communication to assess an effective assembler of communicative resources (amongst other skills), or use an AOP to assess a

student's ability to deliver a prepared and scripted monologue with no use of multimedia. The AOP which involves the use of multimedia would need to be dropped from the assessment suite if discrete linguistic features, or extemporaneous speaking, were considered the target construct.

9.2.2 The Official and Unofficial Tools

The ethnographic data-gathering and interviews shed light on those aspects of practice which were openly shared and those which were applied in a stealth manner. What was openly shared changed depending on who the audience was at a given time. In a room of candid and invested teachers in dialogue, the stealth practice becomes visible.

When reaching scores and considering accountability the criteria was the visible source used in decision making. When reflecting upon the ultimate high-stakes decision and consequential validity of the assessment, teacher assessors exercised critical distance from the criteria at times. Instead, their encounters with students became their evidence of choice in some critical episodes. However, the official line that was toed was that the assessment is mapped to a rating scale and a score is produced. This official line is not immutable: teacher assessors to some extent pay lip service to the construct of the assessment format in language testing but move away from it.

The unofficial practice namely involves shifts from using the AOP performance as the sole source of evidence to opening up the drawbridges to allow judgments made on other sources into the rating discussions. The teacher assessors also turn to assessing the students rather than the AOP performance at times, and compare student performances with each other as a consistency check. The social is brought into the decision making and therefore new validity checks are summoned into use. The complexity of the process with its validity and ethical quandaries is exposed when teachers enter into the unofficial realm.

9.2.3 Importance of Dialogue and Thinking Space

The open discussion in which decisions are co-constructed from seeking reaction from colleagues strengthens the validity of interpretations. Without such space or encouragement, I argue that assessors may fall into routine non-thinking. From reflecting on features of assessment episodes I have encountered in my own practice, I can clearly recount times when assessors have gone through the motions without opportunity to critically reflect on, and alter, decision making. This is

most stark when having to assess high numbers of students based on a short assessment performance in very restricted time frames.

I have shown how teacher assessors can employ a range of tools and enact a series of sophisticated processes in order to enhance the possibility of creating a workable, valid, ethical and ultimately humane assessment process. The implementation, negotiation and evaluation of such tools is time and resource intensive. Provision of the conditions needed to build nuanced interpretations of student ability and opportunities for reflection and dialogue greatly enhances the validity and fairness of the decision making. Students also have the right to voice and justify their own approaches to assessment, particularly in non-standardized assessment. Dedicated time for discussion between assessors and between assessors and test takers after performances would strengthen the validity and fairness of assessment. Such discourse safeguards against teacher assessors and student test takers becoming non-thinking automatons who go through the motions to 'get the assessment done'. As highlighted in Section 5.7, sustained agonized thought is to be cultivated so humane assessment decision making can thrive and be a prominent feature of assessment practice. Research on what may constrain critical inquiry in assessment processes in neoliberal universities would be welcomed.

Dialogue and critical reflection are the core features of the human dimension of assessment practice. The human dimension involves confronting ethical quandaries and adapting practice in light of emerging social influences. Humans can construct both humane and inhumane assessment but only continual critical engagement with decision making from humans can best safeguard against harm.

While the research has gathered information on diverse AOP task formats, purposes, expectations and challenges in S1 and built rich nuanced interpretations of assessment practice in S2, the next section outlines potential limitations of the study.

9.3   Limitations

 The research contains a number of limitations. I first discuss limitations in S1 survey methods and the questionnaire design.

9.3.1   S1 Limitations

The first limitation of the S1 survey phase of this project is that the document analysis and use of questionnaires does not provide data which offers thick description of assessment practice. I endeavoured to yield rich data in the qualitative exploratory questionnaires by using open-ended questions. I acknowledge however that the questionnaire format inevitably will have resulted in brief responses with limited contextual understanding provided. The second limitation in S1 is that the open-ended questionnaire required a high degree of engagement. In future, I would limit the number of questions.

The third limitation is that using qualitative questionnaires is a risky strategy due to the inability to negotiate meaning with the respondents, which is possible in interviews. Interviews were deemed unconducive to the aim of achieving a broad sweep of AOP practice with the resources I had available as the sole investigator on this project. The inability to negotiate meaning in questionnaires, places emphasis on the clarity of question prompts, including choice of terms used. After much deliberation, the terms non-native English speaker and native English speaker were chosen for questions as these were deemed the most recognisable terms to explore perceptions of language use in AOPs. Although this term is open to numerous interpretations with a particular danger of equating native with proficient, alternatives were deemed just as problematic, if not more so. For instance, 'users of English as a first or second language' may not be quickly recognisable to academics across disciplines and created less digestible question prompts. 'Home and international students' was also considered, however this created a deeply uncomfortable inference that nationality shapes English language proficiency. Each term came with serious limitations. I realised any such question makes an assumption that it is clear to SLs which of their students speak English as a first and second language and may signal to respondents that first language users are more proficient users in AOP performance. The survey sought to garner perceptions on any such distinctions between linguistic performances of NESs and NNESs, and did not wish to suggest in any way that there are any. I acknowledge the dangers involved in posing questions in such a way when no negotiation of the meaning of terms is possible and will seek to avoid this in future research.

The survey study (S1) provided information on task formats, purposes, challenges and stated expectations on a large number of subject modules and EAP modules. The Survey Study, however,

provided snapshots of stated practices, the fourth limitation of the S1 survey. The findings of ethnographic research (including in this project) reveals that stated practice diverges from actual practice. The limitations of survey methods have been exposed in the project, however, which is an ultimate strength of this project's overall design which includes an ethnographically-oriented second phase.

9.3.2   S2 Limitations

The first limitation of S2 is that I limited observation to formal classroom and assessment events, except for a number of informal conversations outside class time. For instance, I did not observe students' preparation of the AOP tasks outside of class time: I gathered self-reports of preparation strategies in interviews, instead. I used interviews to access these unobserved teacher and student practices, however it is likely that I did not capture a number of factors which shaped individual practices which I may have in ethnographic observation. I imagine this had ramifications for the capture of the formative dimension of the AOP assessments.

This project investigated teacher and student micro-practices on two EAP courses at two separate institutions. The second limitation in S2 is that I did not gather a wider range of stakeholder perspectives. The third limitation is that the project primarily investigated the micro practices of the individuals engaged in AOP assessment. While macro factors (e.g. university policy, UKVI requirements) were cited by individuals in S1 and S2, I did not engage in a direct analysis of such policies. The fourth limitation is that I conducted research in two settings in S2 which means the theories generated may not be representative of practices on other EAP programmes. The findings may, however, have resonance with other settings, EAP or otherwise (as discussed above in section 4.1).

The fifth limitation is my own experience working and studying in UK HE, namely my experience working as an EAP practitioner, albeit limited. This feature of my background meant that I entered this study with established notions of EAP and AOP assessment. I exercised reflexivity recognising the need to look at practices with fresh eyes. My 'insider status' worked in my favour at times, as I could relate to the practitioners' concerns: in select opportune moments I shared my own challenging experiences with some informants to reassure them that I was aware of some of the complexities and constraints on practice.

Implications for practice and further research have been provided throughout the thesis, but the core suggestions are considered in the next section.

9.4   Implications for Practice and Further Research

The discussion of key findings points to areas which have tangible takeaways for practitioners and lacunae which require intensified research efforts.

The first area with implications for practice and requiring further research is that of raised awareness of communicative demands placed on assessors and test takers in oracy tasks in English-medium institutions. There is a pressing need for a review of the role of communication in assessments which contain an oral component. Not only is there potentially a hidden assessment of language and oracy skills (Doherty et al., 2001; Heron, 2019) but of communication skills more broadly. The act of learning and communicating knowledge is a communicative endeavour. I, therefore, call for ethnographic research which establishes orientations towards English and communication in assessment tasks with an oral component. HE practitioners are strongly encouraged to heighten their awareness of the role of communication in their formative and summative assessment tasks. What communicative demands do the assessment tasks involve? Has there been sufficient instruction on the communication skills which are assessed, or required to fulfil the task? Have practitioners received training in how to implement assessments with complex communicative demands?

The second key implication and stream of research requiring renewed attention is validity enhancing strategies used by stakeholders. Further investigations which expand notions of validity and show what tools and strategies can be, and crucially are, employed by stakeholders is welcome. Practitioners' micro practices offer insights into sophisticated strategies which can be attuned to enhance validity and fairness. I urge stakeholders to consider whether their notions of validity fit with the characteristics of their learning context: do practices done in the name of validity and fairness in fact do harm? Contained within teacher assessor and student test taker practice are flickers of resistance and innovation from which we can learn. Researchers may not need to reinvent the wheel; the resourcefulness of many invested practitioners and students is already grappling with complexities and perplexities of assessment practice. For instance, in this

study the teacher assessors bring the social into the decision making and many students reject the linguistic focus assigned to the task, as communicated in the rubric, and to a smaller extent by the teacher assessors. Teacher assessors, student test takers, assessment developers, programme leaders, external examiners and other stakeholders must continuously question their notions of validity and create space for negotiation. Students and teachers could benefit from workshops dedicated to reflecting on practice as without data reflection can be limited (Kiely and Davies 2010; Slimani-Rolls and Kiely, 2014). The implications for external examiners in the settings which feature in S2, would be that they may need to ask for details of the social context before rendering any actions invalid or unfair. Research on external validation is an area which warrants closer attention, especially when external examiners have little knowledge of the social context. An element of trust needs to be granted to teachers and new forms of audit trails created: the audit and accountability culture of traditional assessment does not map onto non-standardized assessment practice. The teacher assessors in the field sites were experienced and conscious of the ethical dimension of their work. However, the strategies and tools they used remained hidden even to themselves. Another dimension in need of concerted attention is how the interpersonal dimension shapes practice and enhances or threatens validity. Teacher assessors and student test takers would benefit from reflection and research on practice and conscious-raising activities which illuminate their substantive processes. This may prompt teachers and students to confidently and cogently make the case for their thinking and action in assessment practices. What were once unofficial practices tentatively carried out tinged with guilt, if justified, work their way into the official domain.

The third area with implications and need for further investigation relates to student behaviour in AOP assessment. In this study, the central data sets analysed for students in this project output were self-report methods. What the students do in preparing the AOP and in the delivery requires further research, especially their approach to the use of visuals (Kaur and Ali, 2017). Avenues for future research in this sphere of activity may include participant observation of students, or autoethnographies, which document student behaviour during preparation and feedback stages of AOP tasks. Another research focus could involve an analysis of the recorded AOP sessions and rating discussions which examines the treatment of language and communication. The data sets could be subjected to close discursive analysis within an ELF(A) framework. ELF(A) frameworks can offer microanalytic possibilities for the discursive analysis of spoken interaction. I would

caution that such an analysis be complemented by the findings from the content analysis in this study when accounting for communicative decisions taken by teachers and students. Students made the case that the whole communicative act needs to be taken into account. EAP assessments must ensure the rating scales and rater expectations marry up with the realities of student test takers' communicative practice in meeting task requirements. Student test takers' perspectives should be sought and opportunities for students to justify their approaches to assessment tasks granted.

The collation of the expertise of a range of stakeholders is the desired goal in achieving the fairest assessment possible (see Section 2.11 on ethical dimension of assessment). Local assessments would benefit from involving EAP practitioners (with sufficient working understandings of discourse practices in academia), disciplinary tutors and students. Such collaboration and inclusion of a diversity of perspectives is crucial due to the different areas of expertise these stakeholders bring. Disciplinary tutors may have limitations in knowledge as to the discourse requirements involved in producing assessment tasks for instance, and how to develop these skills in others. A key way forward is to create collaboration between students, disciplinary tutors, EAP practitioners and educational and assessment developers. This study shows the complex processes at work in EAP AOP gateway decisions and has valuable contributions to make in our understanding of current and proposed alternatives to standardized language tests.

The implications may apply to a large range of courses in HE. However, S2 findings in particular, may not be relevant to courses in which there is not prolonged and close contact between students and teachers. There is moderate or low engagement with students and teachers on a number of subject courses and EAP courses (namely EAP summer Pre-sessional courses). In terms of validity and fairness, such contact does not engender the construction of a rounded picture of individuals. Further research which offers understanding as to which criteria and strategies can be admitted to bolster validity and fairness in these settings would be welcome.

The research has demonstrated that teacher assessors and student test takers value the integration of linguistic and non-linguistic skills. This has crucial justice implications. Language has been shown to be important in completing the AOP tasks, however overall communicative ability and academic potential seem to have been prioritised. Questions remain as to whether the

second language users on these courses should be the (only) ones to sit these assessments. Dividing students down primarily linguistic and cultural lines may prove to be harmful and unjust. The construct of ability to cope that the teacher assessors are working with seems to engage with this complex issue, moving away from a focus on linguistic features, but this remains unresolved.

9.5   Concluding Remarks

The research process and the findings of this project have, and will, shape my thinking and actions as a researcher and teacher. Researching assessment practices has been an act in developing my assessment (and learning) literacy. In the initial stages of this research, I approached the formal AOPs as a fairly discrete activity, not anticipating the full impact of previous social interactions between the teachers and students and the formative dimension of practice. I have come to better understand the complex sociological processes at work and the agentive roles that teachers and students take on in learning and assessment tasks. The research has made assessment practices visible and discernible to myself and key informants who were not fully aware of what factored into their decision making. The teacher assessors in FS1 and FS2 are heavily influenced by traditional notions of validity although through their behaviour they champion alternative approaches. I was fortunate to have dedicated time and resources to gather insights into AOP tasks from student perspectives, a privilege I did not have in my previous teaching experience. The students shared diverse, justified and often sophisticated approaches to AOP tasks. The value of student and staff dialogue has truly been underscored during this research endeavour.

From taking the time to devote prolonged attention to the art of AOP assessment and learning tasks, the craft involved and the eclectic influences on practice were unearthed. I hope that the findings can be used to raise awareness of the complex mix of factors and strategies involved in preparing, delivering, teaching, assessing, and evaluating AOP assessment tasks. I hope it opens up opportunities for different conceptualisations of validity to gain respectability and to engender greater acceptance of assessment practices which best suit the given learning and social ecology and practices. I have come to strongly believe in the power of ethnographic research in developing grounded and nuanced interpretations of what is happening in learning and assessment practices. I move forward hoping to become gradually more informed and experience enlightening encounters and eye-opening gestalt shifts with my colleagues in the future.

Appendix A Subject Lecturer Questionnaire

1.1 Current Institution

1.2 Faculty

1.3 Discipline

1.4 Number of years' academic teaching experience

1.5 How much experience would you say you have had assessing and giving feedback on academic oral presentations?

2.1 Could you provide an example of an oral presentation task you have assessed at UG or PGT level?

(Please include: level of study, whether it is group/individual, formative/summative, and other key details included on the student brief)

2.2 What components do you assess?

(e.g. presentation and a compulsory supporting handout; a poster and presentation)

2.3 Do you refer to marking criteria/rubrics when assessing and giving feedback on presentations?

○   Yes

○   No

2.4 Please share any comments about the criteria/rubrics and feedback sheets you are provided with (e.g. their usability, appropriateness to task)?

3.1 How are oral presentations a useful assessment task on your module(s)/unit(s)?

3.2 What are the most important features of a successful oral presentation on your taught modules? (i.e. What features would you most commonly prioritise in feedback?)

3.3 In your experience, what aspects of delivering an academic oral presentation do students find most challenging?

Is this similar/different for native and non-native English speakers?

3.4 What kind of allowances (if any) do you make for non-native speakers of English when assessing academic oral presentations?

3.5 Could you provide examples of instances when language use has affected the delivery of a presentation (for native and/or non-native English speakers)? Have you given feedback on these?

(This could include: use of text on slides, register, pace, grammar, word choice, pronunciation)

3.6 What aspects of marking and giving feedback on academic oral presentations have you found challenging?

3.7 Please share any further experiences you have had related to assessing and giving feedback on academic oral presentations

3.8 Please share any further comments, suggestions on the topic or on the questionnaire itself

Appendix B   **EAP Practitioner Questionnaire**

1.1 Current Institution

1.2 Current position (e.g. EAP tutor, EAP course leader)

1.3 Number of years' EAP teaching/administration experience

2.1 Could you provide an example of an oral presentation task you have assessed on an EAP course?

 (Please specify: pre-sessional/in-sessional, group/individual, formative/summative, task details)

2.2 Were students on your course...

- ○ international students only

- ○ home students only

- ○ a mixture of home and international students

- ○ not sure

2.3 Did you refer to marking criteria when assessing and giving feedback on presentations?

- ○ Yes

- ○ No

2.4 What aspects did you assess and what was the weighting for each component?

(e.g. content 25%, language 50%, presentation skills 25%)

2.5 Please share any comments about the criteria and feedback sheets you are provided with (e.g. their usability, limitations)

2.6 Have you been involved in designing oral presentation marking criteria?

○ yes

○ no

2.6b What factors shaped the choice of aspects (e.g. content, fluency, accuracy) assessed?

How did practices on students' degree courses shape the design of criteria (if at all)?

3.1 How are oral presentations a useful assessment task on your EAP course?

3.2 What are the most important features of a successful academic oral presentation on your EAP course? (i.e. What features would you most commonly prioritise in feedback?)

3.3 In your experience, what aspects of delivering an academic oral presentation have your EAP students found most challenging?

3.4 Could you provide examples of instances of language use which may lead to marks being deducted in an EAP oral presentation assessment?

(This could include: use of text on slides, register, pace, grammar, word choice, pronunciation)

3.5 What aspects of marking and giving feedback on academic oral presentations have you found challenging?

3.6 Please share any further experiences you have had related to assessing and giving feedback on academic oral presentations

3.7 Please share any further comments, suggestions on the topic or on the questionnaire itself

Appendix C **Qualitative Questionnaires Rationale and Inspiration**

Question number(s) and topic	Rationale	Main Inspiration
Subject Lecturer Questionnaire		
Qs 1.1-1.5 Institution, Faculty, Discipline, Experience teaching and experience assessing presentations	Gain an understanding of the breadth of the data capture. This was used to gauge whether there was a reasonable degree of diversity in practices represented. Arguably, there is greater capture of diverse practices with respondents from various institutions, disciplines and with varying levels of experience.	Scoping review revealed lack of broad survey of AOP assessment practice in UK HE
Q 2.1 Example of an AOP task	Gather information on different format characteristics Build a more nuanced understanding of what AOP tasks involve and how they are conceptualised	Website survey and documentation showed AOP formats cannot be predicted
Q 2.2 What components are assessed	Consider how AOPs are teamed with other genres in assessment tasks. Gauge how central the AOP task is in the assessment task.	Website survey and documentation revealed AOPs can be teamed with other genres in assessment.
Qs 2.3-2.4 Use of Rubric	Begin to ascertain how rubrics are used and perceived as these are often integral documents in assessment processes, at least in how an institution claims validity.	Website survey and documentation showed rubrics were a core document used in assessment tasks Scoping review showed that rubrics analysis of rubrics can offer critical insights into AOP practice
Q 3.1 Usefulness of AOP task	Gather information on how the AOP aligns with the module objectives and what functions it may perform.	Levrai and Bolster (2015) questions for subject lecturers
Q. 3.2 Most important features	Access what counts for subject lecturers in AOPs. This information could also be used to inform EAP needs analysis.	Levrai and Bolster (2015) questions for subject lecturers
Q.3.5 Examples of feedback on language use	Ascertain whether instances of language use are regarded as warranting enough attention that they feature in feedback on AOPs. This would indicate the importance placed on, and role associated with, language use and communication in AOPs	Doherty et al. (2011) finding that there was a great potential that a 'hidden assessment of oracy' was taking place in subject modules in Australian tertiary contexts.

Qualitative Questionnaires Rationale and Inspiration

Question number(s) and topic	Rationale	Main Inspiration
Q.3.6 Challenges for subject lecturer	This question may point to unknown or unresolved tensions which are important to consider when investigating AOPs.	Challenges are explored in previous studies on teacher and student assessment and offer critical insights into approaches (e.g. Zappa Hollmann 2007a and 2007b)
EAP Practitioner Questionnaire		
Question number(s) and topic	Rationale	Main Inspiration
Qs 1.1-1.3 Institution, position and years' experience	Gain an understanding of the breadth of the data capture. This was used to gauge whether there was a reasonable degree of diversity in practices represented.	Scoping review revealed lack of broad survey of AOP assessment practice on EAP courses, despite a wealth of practice. I had witnessed the wealth of practice from my experience as a practitioner and BALEAP events.
Q. 2.1 Example of a presentation	Gather information on different format characteristics Build a more nuanced understanding of what AOP tasks involve and how they are conceptualised	Website survey and documentation showed AOP formats cannot be predicted and more examples act as points of comparison which offer closer understandings of specific practices.
Q. 2.2 Students on the course (home and/or international)	Determine any division down linguistic lines and potential impact on assessment practice.	Attendance at BALEAP events revealed many institutions now offer courses to home students.
Q.2.3-2.5 Use of rating scale and what criteria included in the scale	Begin to ascertain how rubrics are used and perceived as these are often integral documents in assessment processes, at least in how an institution claims validity	Website survey and documentation showed rubrics were a core document used in assessment tasks Scoping review showed that rubrics are often analysed in assessment studies and provide crucial insights into practice.
Q. 2.6 Design of rating scale	Garner information on possible factors shaping the design of rubric. Ascertain the role of the criterion in assessment design.	Subject lecturers provided insightful comments on rubric design process which indicated this should feature in the EAP Practitioner questionnaire.
Q. Usefulness of AOP tasks on EAP module	Gather information on how the AOP aligns with the module/course objectives and what functions it may perform.	Levrai and Bolster (2015) questions for subject lecturers

Qualitative Questionnaires Rationale and Inspiration

Question number(s) and topic	Rationale	Main Inspiration
Q. 3.2 Most important features of AOP	Access what counts for EAP practitioners in AOPs. This information could also be used to evaluate EAP needs analysis and theorise the construct of EAP.	Levrai and Bolster (2015) questions for subject lecturers
Q. 3.3 Challenges for students	This question may point to unknown or unresolved tensions which are important to consider when investigating AOPs, especially validity issues	Challenges are explored in previous studies on teacher and student assessment (e.g. Zappa Hollmann 2007a and 2007b)and offer crucial insights into practice.
Q. 3.4 Instances of language use led to deduction of marks	This would indicate the importance placed on, and role associated with, language use and communication in AOPs by EAP practitioners.	Scoping review reveals different beliefs as to role and emphasis of second language in performance assessments (e.g. McNamara 1996) and in academic communication (e.g. Canagarajah, 2006)
Q. 3.5 Challenges for teachers	This question may point to unknown or unresolved tensions which are important to consider when investigating AOPs, especially validity issues.	Challenges are explored in previous studies on teacher and student assessment (e.g. Zappa Hollmann 2007a and 2007b) and offer crucial insights into practice.

Appendix D **Subject Lecturer Disciplines**

S1 Subject Lecturer Questionnaire respondents stated they were in the following disciplines.

There were more than one respondent for some of the disciplines listed below.

Subject Lecturer Respondents' Stated Disciplines		
Media	Interdisciplinary Social Scientist	Education
Law	Anthropology	Pharmacy
Modern Languages	Archaeology	Accounting
Zoology	Physics	Sports Medicine
Environmental Science	Genetics	Veterinary Medicine
Medicine	Sociology	Biomedical Engineering
History of Art	Urban Studies	Animal Behaviour and Welfare
Social and Public Policy	Nutrition	Management
Music	Engineering	Applied Linguistics
Psychology	Politics	Economics

Appendix E **Prompts for Interview Before Presentation with Students**

Researcher's and student's prompt of interview topics in interviews with students before the presentation

General discussion about yourself

Lessons and your research project

Experience delivering presentations

What you learnt about delivering presentations in lessons

Your practice presentation

What you think the markers expect in your presentation

The marking criteria – familiar with it? Helpful?

Your presentation preparation (if you have started!)

Challenging aspects and enjoyable aspects of presentation

What language your assessors want to hear and see in your presentation

What specific strategies you and your friends use to cope with challenging aspects

What you enjoy about the course

Anything you would like to know / learn/ have help with / think about to help you deliver an academic presentation

Any comments / experiences / reflections you would like to share

Appendix F Prompts for Interview Before Presentation with Teachers

Researcher's interview prompts for interview with teachers before the presentation

Talk me through… could you tell me more about… what are your thoughts about…

- course details and how similar/different to other courses at institution
- design of the course
- oral presentation format recap
- your and students' views of purpose of the oral presentation
- the degree of influence on the AOP assessment task
- what makes a successful /unsuccessful presentation on the course
- design of the criteria, what factors shaped the design, weighting in the marking criteria, recent changes, consider changing any aspects
- students and criteria
- instruction, aspects covered related to presentations
- role of linguistic and non-linguistic factors in the oral presentation
- instances of language use lead to deduction in mark
- student strengths and challenges
- any experience with L1 students? any differences?
- positives in AOP task
- challenging or ambiguous areas in teaching and assessing AOPs
- further experiences/ comments/anecdotes

Appendix G **Prompts for Interview After Presentation with Students**

Reflection task and Researcher's prompts for interview with students after the presentation

- Topic of presentation
- Who was in your group
- Tell me about the experience

- Questions referring to the reflection sheet if appropriate

- How do you feel about the presentation?

- Talk me through the preparation

- Was the experience what you expected?

- What do you do when presenting?

- What do you notice as an audience member?

- What do you value in the presentations?

- What would you include in the marking criteria?

Appendix H **Prompts For Interview After Presentation with**

Teachers

Researcher's prompts for interview with teachers after marking presentations

Talk me through… could you tell me more about… what are your thoughts about…

- overall thoughts about student performances as a whole.
- what you do when watching presentation
- what you do when agreeing marks with co-assessor
- what goes through your mind
- who are you making judgements as (e.g. you as a teacher, an examiner, an unforgiving interlocutor in target domain…)
- compare presentations in different contexts, by L1 or previous years
- students incorporate input from class
- your use of criteria during the discussion, referral to parts of the video for some students? Clips chosen?
- thoughts on criteria after the marking – any areas which would change?
- specific context shape the approach to this task
- view on the importance of content and delivery in a presentation
- weighting of different aspects
- factors that shape views on content, delivery, presentation skills, teamwork do you think?
- difficulty noticing language use
- role of Q&A
- assessing teamwork
- constraints on what you do
- assessment literacy
- moving forward, what aspects might change about criteria, lesson input, marking process, feedback

Appendix I **S1 NVivo Coding**

Below is a screenshot of part of the codes generated during analysis of S1 EAPQ data. The node featured in this extract pertains to the data analysed most closely related to Question 3.1 (See Appendix B above). The final codes were tweaked in memo writing and when writing up the draft and final report. The analysis related to the nodes featured in this screenshot are discussed in Section 4.3.1.

Nodes

Name	Files	Referen
EAP lecturer questionnaire	0	0
AOP useful	2	56
clear assesment format for students	1	1
delevelops academic skills	2	6
develops group work skills	2	2
develops presentation kills	2	3
develops confidence	2	8
diversifies development of core academic skills	1	1
diversifies assessment task types on EAP programmes	1	1
easier to assess than seminars	2	3
encourage students to be succint	2	2
motivates students	1	2
provides opportunity to practise speaking skills	2	9
provides opportunity to practise listening skills	2	2
authentic task used on subject courses	2	10
formative learning experience	2	2
speaking skills can be assessed	1	4
affords assessment of integrated skills	0	0
affords ability to gauge students' understanding of a topic	0	0
Important AOP features	2	55

Appendix J **S1 Memo Writing**

This is a screenshot of the list of memos written on S1 data (see right). I created a memo for each of the main topics covered in the questionnaire and used the memo to draw together the codes and data from EAPQ, SLQ and documentation analysis.

PhD S1 writing Search Project

Name	Codes	Reference
PhD SL questionnaire further experiences and comme	18	24
PhD SL questionnaire components assessed	7	8
PhD SL questionnaire challenges student	19	48
PhD SL questionnaire analysis language feedback	17	18
PhD SL questionnaire analysis criteria comments	17	24
PhD SL questionnaire analysis challenges teacher	20	23
PhD SL questionnaire analysis AOPs useful	32	77
PhD SL questionnaire analysis AOP important features	13	23
PhD SL questionnaire analysis allowances	22	37
PhD EAP useful task	21	47
PhD EAP student challenges node	12	20
PhD EAP staff challenges	6	9
PhD EAP most important features node writing	22	50
PhD EAP language use tretment	10	16
PhD EAP design of criteria	8	14
PhD EAP criteria comments	4	6
PhD EAP aspects assessed	18	49

Below is screenshot of a memo extract on the questionnaire data and coding on 'Important features of AOPs'. I used the memo to identify the range of positions on the role of content and delivery (which is discussed in 4.3.2)

PhD SL questionnaire analysis A ✕   PhD SL questionnaire analysis   PhD SL questionnaire analysis I

Content and Delivery emphasis

43 lecturers referred to content in some form whereas 23 referred to presentation skills/delivery aspects in their response to which features were more important to an AOP.

A number of responses gave an indication as to the differing emphasis which is placed on content and delivery. The following responses to question show how content may be the priority (subject lecturer 20) delivery may take precedence (subject lecturer 26) or equal emphasis may be placed on content and delivery (subject lecturer 36)

Content is first.
Speaking clearly, slowly, inclusively is second.
Both are essential for an A grade.
Subject lecturer 20, Sociology

Delivery & Timing, followed by content and visual aids.
The second two are easier to improve, but the first two are in some ways totally critical to something being well receieved at a conference, interview, briefing, etc
Subject lecturer 26, Archaelogy

1) The quality and depth of information presented.
1) The speed and clarity of oral and/or visual material
2) Audience engagement.
Subject lecturer 36, Urban Studies

Appendix K **S2 NVivo Coding**

The following screenshot shows a selection of nodes in NVivo. I coded data using the 'nodes' feature on NVivo to generate focused codes and their emerging properties. Further analysis was conducted after this coding (in memos and in writing up) meaning the codes and properties, and their labels in the screenshot, may be different to those in this written report: for example 'Gathering for evidence' became 'Gathering and Selecting Evidence'. This screenshot (below) shows the workings out, but not the final working calculation.

Gathering for evidence	12	292
Critically reflecting	12	287
occupying teacher and examiner spaces	7	21
aligning rating behaviour with current assessment	7	9
making affordances	5	25
considering validity and fairness of decisions	8	51
Balancing freedom and guidance	4	8
acknowledging affective factors	7	20
referring to target domain practices	6	34
reaching justifications for tough assessment decisions	5	13
giving space for experimentation	7	15
using insider knowledge	9	91
Managing Personal relationships	8	95

Appendix L **S2 Memo Writing**

The following screenshot contains part of a memo which was written in order to identify and develop core codes and properties. This memo facilitated establishing the significance of, and the fleshing out of, the code 'The Ranking Filter', which is discussed in Chapter 6. In such memos, I worked to define and describe the characteristics of the actions identified in initial parts of coding. I also compared data extracts in order to ascertain the fit of the code with practices within and across sites.

Comparing students as a marking consistency strategy (23 references) telling code

The raters constantly compare students' performances and marks with other students as a strategy to ensure the consistency of marks. This seemed to be the main strategy used when making decisions on marks for sections and then with the final allocation of overall marks for the students. The raters in parts of the rating discussions stated how the mark may be the same but for different reasons, but the overall perforamance was on the same grading level as another student. Simply looking at the criteira and not comaparing to other students did not seem to fill the raters with enough confidence in the consistency of their marking.

Ref 1 - reference about presentation skills section mark comparision

Adam:
Ok. Mm-hm. Um well 12 presentation skills? Or 13 the same as Henry.
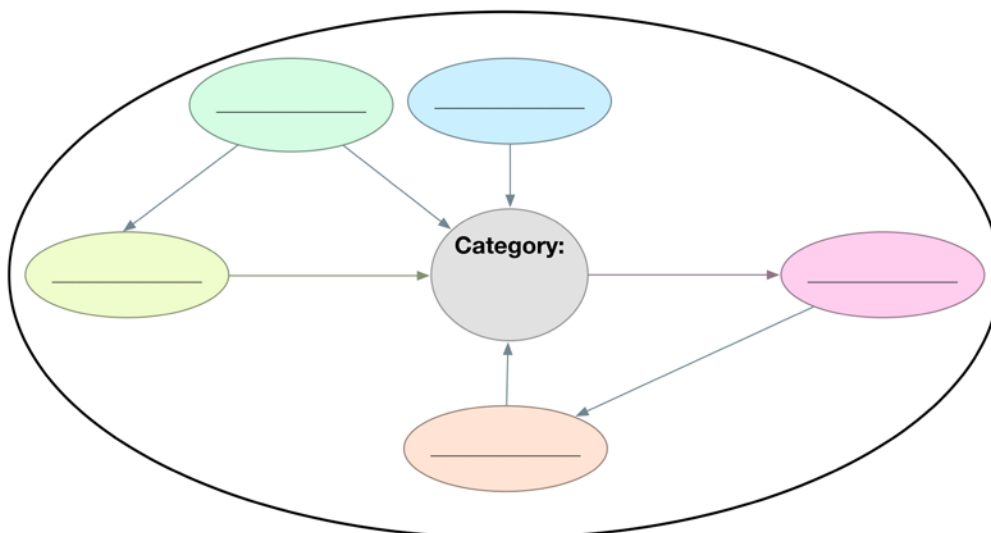
Daniel:
She wasn't as good as Henry.

Adam:
12 then [or 11?]

Appendix M   **Theoretical Coding Worksheet**

Created by Professor Gregory Hadley and expanded upon in Hadley (2019)

Theoretical Coding Worksheet

Write the category (or concept) created from your focused codes in graph below.

Choose from the following concepts (Böhm 2004, Glaser 1973) below and place them in the graph as questions for further exploring the way that your category works and operates.

Coding families	Concepts	Examples
The Six Cs	Causes, contexts, contingencies, consequences, conditions	… of pain suffering
Process	Stages, phases, phasings, transitions, passages, careers, chains, sequences	Career of a patient with chronic pain
The Degree Family	Extent, level, intensity, range, amount, continuum, statistical average, standard deviation	Extent of pain suffering
Type Family	Types, classes, genres, prototypes, styles, kinds	Kinds of pain – sharp, piercing, throbbing, shooting, sting, gnawing, burning
The Strategy Family	Strategies, tactics, techniques, mechanisms, management	Coping with pain
Interactive Family	Interaction, mutual effects, interdependence, reciprocity, symmetries, rituals	Interaction of pain experience and coping
Identity-Self Family	Identity, self-image, self-concept, self-evaluation, social worth, transformations of self	Self-concepts of pain patients
Cutting-Point Family	Boundary, critical juncture, cutting point, turning point, tolerance levels, point of no return	Start of chronification in the medical career of pain patient
Cultural Family	Social norms, social values, social beliefs	Social norms about tolerating pain, 'feeling rules'
Consensus Family	Contracts, agreements, definitions of the situation, uniformity, conformity, conflict	Compliance

Go into the field to find answers to those questions that you do not know. Search for scholarly literature as well. Write memos that explain your category along the lines of the concepts. Then ask yourself, at the end 'What else must be true in order for your explanation to be true?' This will help you to think about further dynamics and issues in your grounded theory.

List of References

Adler, P. A., and Adler P. (1998) 'Observational techniques', in Denzin, N. K. and Lincoln, Y. S. (eds) *Collecting and interpreting qualitative materials.* Thousand Oaks, CA: Sage publications, pp.377-392.

Airey, J. (2012) ''I Don't Teach Language': The Linguistic Attitudes of Physics Lecturers in Sweden.' *AILA Review 25*, pp. 64–79.

Alderson, J. C. (1980) 'Native and non-native speaker performance on cloze tests', *Language Learning*, 30, pp. 59-76.

Arendt, H. (1955) *Men in dark times*. Middlesex, UK: Penguin Books.

Atkinson, P. and Silverman, D. (1997) 'Kundera's Immortality: The interview society and the invention of the self', *Qualitative Inquiry, 3*, pp. 304-325.

Austin, J. L. (1962) *How To Do Things With Words.* 2nd edn. Edited by J.O. Urmson and M. Sbisá. Cambridge, MA: Harvard University Press.

Bachman, L. F. (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. and Palmer, A. S. (1996) *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L. F. (2007) 'What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment', in Fox J., Wesche, M., Bayliss, D., Cheng, L., Turner, C. and Doe, C. (eds) *Language testing reconsidered*. Ontario: University of Ottawa Press. pp. 41–71.

Bachman, L. F. and Palmer, A. S. (2010) *Language Assessment in practice*. Oxford: Oxford University Press.

Baird, R., Baker, W., Kitazawa, M. (2014) 'The Complexity of ELF', *Journal of English as a Lingua Franca,* 3 (1), pp. 171-196.

Baker, W. (2015) *Culture and Identity through English as a Lingua Franca*. Berlin: De Gruyter Mouton.

Baker, W. and Hüttner, J. (2019) '"We are not the language police": Comparing multilingual EMI programmes in Europe and Asia', *International Journal of Applied Linguistics,* 29 (1), pp. 78-94.

Bamkin, M., Maynard, S. and Goulding, A. (2016) 'Grounded theory and ethnography combined; A methodology to study children's interactions on children's mobile libraries', *Journal of Documentations,* 72 (2), pp. 214-231.

Bankowski, E. (2010) 'Developing skills for effective academic presentations in EAP', *International Journal of Teaching and Learning in Higher Education*, 22, pp. 187–196.

Barrett, N. E. and Liu, G.Z. (2016) 'Global Trends and Research Aims for English Academic Oral Presentations: Changes, Challenges, and Opportunities for Learning Technology', *Review of Educational Research*, pp. 1-45.

Bassey, M. (1996) 'Three Paradigms in Educational Research', in Pollard, A. (ed.) *Readings for Reflective Teaching in Primary Schools*. London: Cassell, pp. 43-5.

Battersby, D. (1981) 'The use of ethnography and grounded theory in educational research', *McGill Journal of Education,* 16 (1), pp. 91-98.

Belcher, D. (2006) 'English for specific purposes: teaching to perceived needs and imagined futures in worlds of work, study, and everyday life', *TESOL Quarterly,* 40, pp. 133–156.

Bell, M. L. (2010) *Interlocutor-candidate interaction in the question phase of academic oral presentations: a validation study*. PhD thesis. University of Bristol.

Benesch, S. (2001) *Critical English for Academic Purposes: Theory, Politics, and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bennett, R. (2011) 'Formative assessment: a critical review', *Assessment in Education: Principles, Policy and Practice',* 18, (1), pp. 5-25.

Biggs, J. (2003) *Teaching for quality learning at university*, Maidenhead: SRHE.

Blommaert, J. (2010) *The sociolinguistics of globalization*. Cambridge: Cambridge University Press.

Blommaert J. and Dong, J. (2010) *Ethnographic Fieldwork: A beginner's guide*. Bristol: Multilingual Matters.

Bogdan, R. C., and Bilken, S. K. (2007) *Qualitative Research for Education*. Boston: Pearson Education.

Böhm, A. (2004) 'Theoretical Coding: Text Analysis in Grounded Theory, In U. Flick, E. Kardorff & I. Steinke' (Eds.) A Companion to Qualitative Research. London: Sage Publications. pp. 270-275.

Bonner, S. M. (2013) 'Validity in classroom assessment: purpose, properties and principles', in McMillan, J. H. (ed.) *Sage handbook of research on classroom assessment*. Thousand Oaks, CA: Sage, pp. 87-106.

Bourdieu, P. (1991) *Language and Symbolic Power*. Cambridge: Polity Press.

Boyle, R. (1996) 'Modelling oral presentations', *ELT Journal*, *50* (2), pp. 115-126.

Breiteneder, A. (2005) 'The naturalness of English as a European lingua franca: The case of the 'third person -s'*, Vienna English Working Papers, 14 (2), pp. 3–26*.

Brinberg, D. and McGrath, J. E. (1985) *Validity and the research process*. Newbury Park, CA: Sage.

Brockmann, M. (2011) 'Problematising short-term participant observation and multi-method ethnographic studies', *Ethnography and Education*, 6 (2), pp. 229-243.

Brookhart, S. M. (2003) 'Developing measurement theory for classroom assessment purposes and Use', *Educational Measurement: Issues and Practice, 22* (4), pp. 5–12.

Brown, A. (1995) 'The effect of rater variables in the development of an occupation-specific language performance test', *Language testing,* 12 (1), pp. 1-15.

Bruce, I. (2017) 'Knowledge and the EAP Practitioner' *The 2017 BALEAP RES TES Symposium: Knowledge and the EAP Practitioner*, University of Leeds, 28 January 2017. pp. 1-8. Available at: https://teachingeap.files.wordpress.com/2017/01/ian-bruce-baleap-restes-paper-january-28.pdf (Accessed 25 January 2017).

Bryant, A. and Charmaz, K. (2007) 'Grounded Theory Research: Methods and Practices', in Bryant, A. and Charmaz, K. (eds) *The Sage Handbook of Grounded Theory*. London: Sage, pp. 1-28.

Burrel, G. and Morgan, G. (1979) *Sociological paradigms and organizational analysis*. London: Heinemann Educational Books.

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

Butler, J. (2007) *Gender Trouble: Feminism and the Subversion of Identity*. London: Routledge.

Cain, S. (2012) *Quiet: The power of introverts in a world that can't stop talking*. London: Penguin Books.

Canagarajah, A. S. (2006) 'Changing communicative needs, revised assessment objectives: Testing English as an international language?', *Language Assessment Quarterly,* 3 (3), pp. 229-242.

Cananarajah, S. (2018) 'Translingual Practice as Spatial Repertoires: Expanding the Paradigm beyond Structuralist Orientations' *Applied Linguistics*, 39 (1), pp.31–54.

Chalhoub-Deville, M. (2003) 'Second-language interaction: Current perspectives and future trends', *Language Testing*, 20 (4), pp. 369-83.

Chandler, D. (1997) 'An Introduction to Genre Theory', Available at: http://visual-memory.co.uk/daniel/Documents/intgenre/chandler_genre_theory.pdf (accessed July 2019)

Chapelle, C. A. (2012) 'Conceptions of validity', in Fulcher, G. and Davidson, F. (eds) *Routledge Handbook of Language Testing.* London: Routledge, pp. 21-33.

Chapelle, C. A. (1998) 'Construct definition and validity inquiry in SLA research', in Bachman, L. F. and Cohen, A. D. (eds) *Inferfaces between Second Language Acquisition and Language Testing Research*. New York: Cambridge University Press, pp. 32-70.

Charmaz, K. (2000) 'Grounded Theory: Objectivist and Constructivist Methods', in Denzin, N. and Lincoln, Y. (eds) *Sage Handbook of Qualitative Research*. 2nd edn. Thousand Oaks: California: Sage publications, pp. 509-535.

Charmaz, K. (2006) C*onstructing Grounded Theory: A Practical Guide through Qualitative Analysis*. 1st edn. London: Sage Publications.

Charmaz, K. (2013) Interviewed by Graham Gibbs at BPS Qualitative Methods in Psychology Section Conference University of Huddersfield, 4-6 September. Available at: https://www.youtube.com/watch?v=D5AHmHQS6WQ (accessed March 2017)

Charmaz, K. (2014) *Constructing grounded theory*. 2nd edn. Thousand Oaks, CA: Sage publications.

Charmaz, K. (2018) *Introduction to Constructivist Grounded Theory* [Workshop]. Lancaster University. July 2018.

Charmaz, K. and Mitchell, R. G. (2001) 'Grounded Theory in Ethnography', in Atkinson P., Coffey A., Delamont, S., Lofland, J. and Lofland, L. (eds) *Handbook of Ethnography.* London: Sage publications, pp. 160-174.

Chopin , K. (2014) 'Reconceptualising norms for language testing: Assessing English language proficiency from within an ELF framework', in Bayyurt T. Y. and Sumru, A. (eds) *Current perspectives on pedagogy for English as a Lingua franca.* Berlin: De Gruter Mouton, pp.193-204.

Christians, C. (2005) 'Ethics and politics in qualitative research', in Denzin, N. K. and Lincoln, Y. S. (eds) *The sage handbook of qualitative research.* 3rd edn. London: Sage publications, pp. 139-164.

Cohen, L., Manion, L. and Morrison, K. (2007) *Research Methods in Education*. 6th edn. London: Routledge.

Colley, K. (2008) 'Performance-based assessment', *Science Teacher*, 75 (8), pp. 68-72.

Connell, J., Lynch, C. and Waring, P. (2001) 'Constraints, Compromises and Choice: Comparing Three Qualitative Research Studies', *The Qualitative Report,* 6 (4). Available at: https://nsuworks.nova.edu/tqr/vol6/iss4/1/

Connor, U. (2004) 'Intercultural Rhetoric Research: Beyond Texts', *Journal of English for Academic Purposes*, 3 (4), pp. 291-304.

Creswell, J. W. (2013) *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage Publications.

Davies, A. (1992) 'Is language proficiency always an achievement?', *Melbourne Papers in Language Testing*, 1 (19), pp. 1-17.

Denzin, N. K. and Lincoln, Y. S. (2003) *Collecting and Interpreting Qualitative Materials.* Thousand Oaks California: Sage.

Derrida, J. (1981) 'The law of genre', in Mitchell, W. J. T. (ed.) *On Narrative*. Chicago: University of Chicago Press.

Devitt, A. J. (1993) 'Generalizing About Genre: New Conceptions of an Old Concept', *College Composition and Communication,* 44 (4), pp. 573–86.

Dey, I. (1999) *Grounding Grounded Theory*. San Francisco, CA: Academic Press.

Ding, A. (2016) 'Neoliberal EAP: are we all neoliberals now?', *Teaching EAP, 25 January,* Available at: https://teachingeap.wordpress.com/2016/01/25/neo-liberal-eap-are-we-all-neoliberals-now/ (Accessed January 2017)

Ding, A., and Bruce, I. (2017). *The EAP practitioner: Operating on the edge of the academy*. Basingstoke, England: Palgrave.

Dippold, D., Bridges, S., Eccles, S., and Mullen, E. (2019) 'Taking ELF off the shelf: Developing HE students' speaking skills through a focus on English as a lingua franca', *Linguistics and Education*, 54, pp. 1-10.

Doherty, C., Kettle, M., May, L. and Caukill, E. (2011) 'Talking the talk: oracy demands in first year university assessment tasks', *Assessment in Education: Principles, Policy & Practice,* 18 (1), pp. 27-39.

Dornyei, Z. (2007) *Research methods in Applied Linguistics*, Oxford, Oxford University Press.

Duff, P. A. (2010). 'Language Socialization into Academic Discourse Communities', *Annual Review of Applied Linguistics*, 30, pp. 169-192.

Dunn, K. E. and Mulvenon, S. W. (2009) 'A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education', *Practical Assessment Research & Evaluation*, 14 (7), pp. 1-11.

Dunning, D. and Krueger, J. (2000) 'Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments', *Journal of Personality and Social Psychology*, 77 (6), pp.1121-34

Elder, C. (2016) 'Exploring the limits of authenticity in LSP testing', *Language Testing*, 33 (2), pp. 1-6.

Ferris, D. (1998) 'Students' views of academic aural/oral skills: A comparative needs analysis', *TESOL Quarterly,* 2, pp. 289–318.

Ferris, D., and Tagg, T. (1996a) 'Academic oral communication needs of EAP learners: What subject-matter instructors actually require', *TESOL Quarterly,* 30 (1), pp. 31–58.

Ferris, D, and Tagg, T. (1996b) 'Academic listening/speaking tasks for ESL students: Problems, suggestions, and implications', *TESOL Quarterly,* 30 (1), pp. 297–317.

Flower, L., Stein, V., Ackerman, J., Kantz, M. J., McCormick, K., and Peck, W. (1990) *Reading to write: Exploring a cognitive and social process*. New York: Oxford University Press.

Flowerdew, J., and Peacock, M. (2001) *Research Perspectives on English for Academic Purposes.* Cambridge: Cambridge University Press.

Foucault, M. (1979) *Discipline and punish*. New York: Vintage Books.

Fulcher, G. (2012) 'Scoring performance tests', in Fulcher, G. and Davidson, F. (eds) *The Routledge Handbook of Language Testing*. London: Routledge, pp. 378-392.

Fulcher, G. and Davidson, F. (2007) *Language testing and assessment: an advanced resource book*. Abingdon: Routledge.

Garner, M. and Borg, E. (2004) 'An ecological perspective on content-based instruction', *Journal of English for Academic Purposes*, 4 (2), pp. 119-134.

Gee, J. P. (1996) *Social Linguistics and Literacies*. 2nd edn. London: Taylor and Francis.

Ginkel,V. S., Laurentzen, R., Mulder, M., Mononen, A., Kyttä, J. and Kortelainen, M. J. (2017) 'Assessing oral presentation performance: Designing a rubric and testing its validity with an expert group', *Journal of Applied Research in Higher Education*, 9 (3), pp. 474-486.

Gillet, A. (2013) 'The Relationship between Academic Content and Language' *UEFAP*, 11th July, Available at: http://www.uefap.net/blog/?p=444 (accessed 21/04/2020).

Gipps, C., Brown, M., McCallum, B. and McAlister, S. (1995) *Intuition or Evidence?* Buckingham: Open University Press.

Glaser B.G. and Strauss A. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine De Gruyter, New York.

Glisan, E., Uribe, D., & Adair-Hauck, B. (2007) 'Research on integrated performance assessment at the post-secondary level: Student performance across the modes of communication', *Canadian Modern Language Review*, 64 (1), pp. 39-67.

Goffman, E. (1967) *Interaction ritual*. New York: Doubleday.

Graham, J. G. (1987) 'English language proficiency and the prediction of academic success', *TESOL quarterly*, 21 (3), pp. 505-21.

Green, A. (2014) *Exploring language assessment and testing: Language in action*. Abingdon: Routledge.

Hadley, G. (2012) *Blended EAP professionals in corporatized higher educational institutions: a critical grounded theory.* PhD thesis. University of Birmingham.

Hadley, G. (2015) *English for Academic Purposes in Neoliberal Universities*: *A Critical Grounded Theory*, New York: Springer.

Hadley, G. (2017) *Grounded Theory in Applied Linguistics Research: A practical guide.* Abingdon: Routledge .

Hadley, G. (2019) 'Critical Grounded Theory' in Bryant, T. & Charmaz, K. (eds) *The Sage Handbook of Current Developments in Grounded Theory*. London: Sage publications, pp. 564-592

Hall, C. (2014) 'Moving beyond tests of English to tests of 'Englishing'', *ELT Journal*, 48 (4), pp. 376–385.

Hamilton, J., Lopes, M., McNamara, T. and Sheridan, E. (1993) 'Native-speaker performance on Tests of English for Academic Purposes', *Language Testing*, 10, pp. 337–353.

Hammersley, M. and Atkinson, P. (2007) *Ethnography: Principles in Practice.* 3rd edn. London: Routledge.

Hammersley, M. and Atkinson, P. (2019) *Ethnography: Principles in Practice.* 4th edn. London: Routledge.

Hamp-Lyons, L. (1997) 'Ethics in language testing', in Clapham, C. and Corson, D. (eds) *Encyclopedia of language and education*, vol. 7, *Language Testing and Assessment*. Dordrect: Kluwer Academic Publishers, pp. 32-333.

Hamp-Lyons, L. (2015) **'***The future of JEAP and EAP', Journal of English for Academic Purposes,* 20, pp. A1–A4.

Harding, L. and McNamara, T. (2018) 'Language assessment: The challenge of ELF', in Jenkins, J. Baker, W., and Dewey, M. (eds) *The Routledge handbook of English as a Lingua Franca*. London: Routledge, pp. 570-582.

Harwood, N. and Petrić, B. (2011) 'English for academic purposes' in Simpson J. (ed.), *Routledge handbook of applied linguistics.* London: Routledge, pp. 243-258.

Hawkins, M. (2018) 'Transmodalities and Transnational Encounters: Fostering Critical Cosmopolitan Relations', *Applied Linguistics,* 39 (1), pp. 55–77.

Hayward, T. (1995) *Ecological thought: An introduction*. Cambridge: Polity Press.

Hill, M. (2009) 'Ways of seeing: using ethnography and Foucault's 'toolkit' to view assessment practices differently', *Qualitative research,* 9 (3), pp. 309-330.

Hill, K. and McNamara, T. (2011) 'Developing a comprehensive, empirically-based research framework for classroom-based assessment', *Language Testing,* 29 (3), pp. 395-420.

Hill, M. and Storey, A. (2003) 'SpeakEasy: online support for oral presentation skills', *ELT Journal*, 57 (4), pp. 370-376.

Heron, M. (2019) 'Making the case for oracy skills in higher education: practices and opportunities', *Journal of University Teaching and Learning Practice*, 16 (2), pp. 1-16.

Holliday, A. (2007) *Doing and writing qualitative research*. 2nd edn. London: Sage publications.

Hounsell, D., N. Falchikov, J. Hounsell, M. Klampfleitner, M. Huxham, K. Thompson, and S. Blair. (2007) *Innovative assessment across the disciplines: An analytical review of the literature*. York: Higher Education Academy. Available at http://bit.ly/2QlmcHi

Hughes, J. (1990) *The Philosophy of Social Research.* 2nd edn. Longman: London.

Hultman, K., and Lenz Taguchi, H. (2010) 'Challenging anthropocentric analysis of visual data: A relational materialist methodological approach to educational research', *International Journal of Qualitative Studies in Education, 23*, pp. 525–542.

Huxham, M., Campbell, F. and Westwood, J. (2012) 'Oral versus written assessments: a test of student performance and attitudes', *Assessment & Evaluation in Higher Education*, 37 (1), pp. 125-136.

Hyland, K. (2006) *English for Academic Purposes: An Advanced Resource Book*. London: Routledge

Hyland, K. (2009) *Academic discourse: English in a global context*, Continuum International Publishing Group.

Hyland, K. (2018)' Sympathy for the devil', *Language Teaching*. 51 (3), pp. 383-399.

Hyland, K. and Shaw, P. (2016) 'Introduction', in Hyland, K. and Shaw, P. (eds) *Routledge Handbook of English for Academic Purposes*. London: Routledge, pp. 1-15.

Hymes, D. H. (1972) 'On Communicative Competence', in Pride, J. B. and Holmes, J. (eds) *Sociolinguistics, Selected Readings*. Harmondsworth: Penguin, pp. 260-293.

Januin, J. and Stephen, J. (2015) 'Exploring Discourse Competence Elements in EAP Class Presentations through Document and Ethnographic Analyses', *Social & behavioural sciences,* 208, pp. 157-166.

Jeffrey, B. and Troman, G. (2004) 'Time for ethnography', *British Educational Research Journal*, 30 (4), pp. 535–548.

Jenkins, J. (2000) *The phonology of English as an international language*: New models, new norms, new goals. Oxford: Oxford University Press.

Jenkins, J. (2014). *English as a Lingua Franca in the international university. The politics of academic English language policy*. London: Routledge.

Jenkins, J. Cogo, A. and Dewey, M. (2011) 'Review of developments in research into English as a lingua franca', *Language Teaching*, 44 (3), pp. 281–315.

Jenkins, J. (2015) 'Repositioning English and multilingualism in English as a lingua franca', *Englishes in Practice,* 2 (3), pp. 49-85.

Jenkins, J. and Leung, C. (2016) 'Assessing English as a Lingua Franca', in Shohamy, E. (ed.) *Encyclopaedia of language education. Language testing and assessment*. vol. 7, New York: Springer, pp.1-15.

Jenkins, J. and Leung, C. (2019) 'From mythical 'standard' to standard reality: the need for alternatives to standardized English language tests', *Language Teaching*, 52 (1), pp. 86-110.

Jenkins, J. and Mauranen, A. (eds) (2019) *Linguistic Diversity on the EMI Campus: Insider accounts of the use of English and other languages in universities within Asia, Australasia and Europe*. Abingdon: Routledge.

Jewitt, C. (2017) 'An introduction to multimodality' in Jewitt, C. (ed.) *The Routledge Handbook of Multimodal Analysis*. Abingdon: Routledge, pp.15-30.

Johnston, P. H. (1989) 'Constructive evaluation and the improvement of teaching and learning', *Teachers College Record, 90 (4), pp.* 509-528.

Jordan, R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.

Joughin, G. (2007) 'Student conceptions of oral assessment in oral presentations', *Studies in Higher Education,* 32, 3, pp. 323-336.

Kalthoff, H. (2013) 'Practices of grading: an ethnographic study of educational assessment', *Ethnography and Education*, 8 (1), pp. 89-104.

Kane, M. T. (1992) 'An argument-based approach to validity', *Psychological bulletin,* 112 (3), pp. 527-535.

Kane, M. T. (2006) 'Validation', in Brennan, R. L. (ed) *Educational measurement*. 4th edn. Westport, CT: Praeger, pp. 17–64.

Kane, M. T. (2013) 'Validating the interpretations and uses of test scores'*, Journal of Educational Measurement*, 50 (1), pp. 1-73.

Kaur, K. and Ali, A. M. (2017) 'Exploring the genre of the academic oral presentations: A critical review', International Review of Applied Linguistics and English Literature, 7 (1), pp. 152-162.

Kiely, R. (2001) 'Classroom Evaluation – values, interests and teacher development', *Language Teaching Research,* 5, pp. 241-261.

Kiely, R., and Davies, M. (2010) 'From transmission to transformation: Teacher learning in English for speakers of other languages', *Language Teaching Research, 14*, pp. 277–295.

Kim, H. and Elder, C. (2009) 'Understanding aviation English as a lingua franca: Perceptions of Korean aviation personnel', *Australian Review of Applied Linguistics*, 32 (3), pp. 23.1–23.17.

Kress, G. (2010) *Multimodality: A social semiotic approach to contemporary communication*. London: Routledge.

Kress, G. (2017) 'What is a mode?', in Jewitt, C. (ed.) *The Routledge Handbook of Multimodal Analysis*. Abingdon: Routledge, pp. 60-75.

Kuby, C. (2017) 'Why a paradigm shift of 'more than human ontologies' is needed: putting to work poststructural and posthuman theories in writers' studio', *International Journal of Qualitative Studies in Education*, 30 (9), pp. 877-896.

Kuby, C., Gutshall Rucker, T. and Kirchhofer, J. M. (2015) ''Go be a writer!': Intra-activity with materials, time and space in literacy learning', *Journal of Early Childhood Literacy*, 15 (3), pp. 394–419.

Labov, W. (1970). The study of language in its social context. Studium Generale, Vol. 23. Reprinted in Labov, W. (1972). Sociolinguistic patterns. Philadelphia, PA: University of Pennsylvania Press.

Langan, A. M., Shuker, D. M., Cullen, W. R., Penny, D., Preziosi, R. F., and Wheater, C. P. (2008) 'Relationships between student characteristics and self, peer, and tutor evaluations of oral presentations', *Assessment and Evaluation in Higher Education,* 33 (2), pp. 179-190.

Lanther, P. (2006) 'Paradigm proliferation as a good thing to think with: teaching research in education as wild profusion', *International Journal of Qualitative Studies in Education*. 19 (1), pp. 35-57.

Lea, M. and Street, B. (1998) 'Student writing in higher education: an academic literacies approach', *Studies in Higher Education*, 11 (3), pp. 182–199.

Lea, M. R. and Street, B. V. (2006) 'The "academic literacies" model: Theory and applications', *Theory into Practice*, 45 (4), pp. 368–377.

Leung, C. (2005) 'Classroom teacher assessment of second language development: Construct as practice' in Hinkel, E. (ed) *Handbook of research in second language teaching and learning.* Mahwah, NJ: Erlbaum, pp. 869–888.

Leung, C. (2007) 'Dynamic assessment: Assessment for and as teaching?', *Language Assessment Quarterly*, 4 (3), pp.257-278.

Leung, C., Lewkowicz, J. and Jenkins, J. (2016) 'English for Academic Purposes: a need for remodelling', *Englishes in Practice*, 3 (3), pp. 55-73.

Leung C. and Teasdale, A. (1997) 'Raters' understanding of rating scales as abstracted concept and as instruments for decision making', *Melbourne Papers in Language Testing*, 6, pp. 45-70.

Levrai, P. and Bolster, A. (2015) 'Developing a closer understanding of Academic Oral Presentations', *Folio,* 16 (2), pp. 65-72.

Li, W. (2018) 'Translanguaging as a practical theory of language', *Applied Linguistics*, 39 (1) pp. 9–30

Lillis, T. and Tuck, J. (2016) 'Academic Literacies: a critical lens on writing and reading in the academy', in Hyland, K. and Shaw, P. (eds) *The Routledge Handbook of English for Academic Purposes.* London: Routledge, pp. 30–43.

Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic Inquiry.* London: Sage publications.

Linn, R. E., Baker, E. L. and Dunbar, S. B. (1991) 'Complex, performance-based assessment: Expectations and validation criteria', *Educational Researcher, 20* (8), pp. 15-21.

Llosa, L. (2008) 'Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments', *Educational Measurement: Issues and Practice*, 27 (3), pp. 32–42.

Lomborg, K. and Kirkevold, M. (2003) 'Truth and validity in grounded theory – a reconsidered realist interpretation of the criteria: fit, work, relevance and modifiability', *Nursing Philosophy*, 4, pp. 189–200.

Macaro, E., Curle, S., Pun, J., An, J. and Dearden, J. (2018) 'A systematic review of English medium instruction in HE', *Language Teaching*, 51 (1), pp. 36-76.

Macfarlane, B. (2016) 'The performative turn in the assessment of student learning: a rights perspective', *Teaching in Higher Education*, 21 (7), pp. 839-853.

Margić, B. D. (2017) 'Communication courtesy or condescension? Linguistic accommodation of native to non-native speakers of English', *Journal of English as a Lingua Franca,* 6 (1), pp. 29–55.

Mauranen, A. (2006) 'Signalling and preventing misunderstanding in English as lingua franca communication', *International Journal of the Sociology of Language*, 177, pp. 123-150.

Maxwell, J. A. (1992) 'Understanding and validity in qualitative research', *Harvard Educational Review*, 62 (3), 279-299.

McNamara, T. (1996) *Measuring Second Language Performance*. Harlow, UK: Longman.

McNamara, T. (1997) '''Interaction" in second language performance assessment: Whose performance?*', Applied Linguistics,* 18, pp. 446–66.

McNamara, T. (2001) 'Rethinking alternative assessment'*, Language Testing,* 18 (4), pp. 329–332.

McNamara, T. (2014) 'Thirty years on – Evolution or revolution', *Language Assessment Quarterly*, 11 (2), pp. 226–232.

McNamara, T. and Ryan, K. (2011) 'Fairness versus Justice in Language Testing: The Place of English Literacy in the Australian Citizenship Test', *Language Assessment Quarterly, 8*, pp. 161-78.

McNamara, T., Knoch, U. and Fan, J. (2019) *Fairness, Justice, and Language Assessment*. Oxford: Oxford University press.

McNess, E., Lore, A. and Crossley, M. (2016) ''Ethnographic Dazzle' and the construction of the 'other': shifting boundaries between the insider and the outsider' in (Eds) Crossley, M. Arthur, L. and McNess, E. *Revisiting Insider-Outsider Research in Comparative and International Education*. Oxford: Symposium Books. pp.21-38.

Mercer, N., Warwick, P. and Ahmed, A. (2017) 'An oracy assessment toolkit: Linking research and development in the assessment of students' spoken language skills at age 11-12', *Learning and Instruction,* 48, pp. 51-60.

Messick, S. (1989) 'Validity' In Linn, R. L. (Ed.) *The American Council on Education/Macmillan series on higher education. Educational measurement* . Macmillan Publishing Co, Inc; American Council on Education, pp.13–103.

Messick, S. (1994) 'The interplay of evidence and consequences in the validation of performance assessments', *Educational researcher*, 23 (2), pp.13-23.

Michell, J. (2009) 'Invalidity in validity', in Lissitz, R. W. (ed) *The concept of validity: revisions, new directions and applications*. Charlotte, NC: Information Age Publishing, p.111-33.

Miles, M., B. and Huberman, A. M. (1994) *Qualitative Data Analysis: an expanded sourcebook*. 2nd edn. Thousand oaks, CA: Sage publications.

Miller, C. R. (1984) 'Genre as social action', *Quarterly Journal of Speech,* 70, pp. 151-67

Mishler, E. G. (1990) 'Validation in inquiry-guided research: the role of exemplars in narrative studies', *Harvard Educational Review*, 60, pp. 415-442.

Morell, T. (2015) 'International conference paper presentations: a multimodal analysis to determine effectiveness', *English for specific purposes*, 37, pp. 137-150.

Morita, N. (2000) 'Discourse socialization through oral classroom activities in a TESOL graduate program', *TESOL Quarterly,* 34, pp. 279–310.

Moss, P. A. (1992) 'Shifting conceptions of validity in educational measurement: Implications for performance assessment', *Review of Educational Research*, 62 (3), pp. 229-258.

Moss, P. A. (1996) 'Enlarging the dialogue in educational measurement: Voices from interpretive research traditions', *Educational Researcher,* 25, pp. 20–28.

Moss, P. A. (2003) 'Reconceptualizing validity for classroom assessment', *Educational Measurement: Issues and Practice, 22* (4), pp. 13-25.

Munby, J. (1978) *Communicative syllabus design*. Cambridge: Cambridge University Press.

Murphy, E. and Dingwall, R. (2003) *Qualitative Methods and Health Policy Research*. New York: Aldine de Gruyter.

Nesi, H. and Gardner, S. (2012*) Genres Across Disciplines*. Cambridge: Cambridge University Press.

Newton, P. and Shaw, S. (2014) *Validity in educational and psychological assessment*. London: Sage. pp.135-181.

North, B. (2000) 'Linking language assessments: an example in a low-stakes context', *System*, 28, pp. 555-77.

Penny, D., Oldekop, P. J. A., Ashkroft, C., Lockey, L., Preziosil, F., Shaw, E. M. and Boyle, J. (2005) 'Peer assessment of oral presentations: Effects of student gender, university affiliation, and participation in the development of assessment criteria', *Assessment and Evaluation in Higher Education*, 30 (1), pp. 21-35.

Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. Mahwah, NJ: Lawrence Erlbaum.

Pennycook, A. (2004) 'Performativity and Language Studies', *Critical Inquiry In Language Studies: An International Journal*, 1 (1), pp. 1-19.

Pennycook, A. (2018) *Posthumanist Applied Linguistics*. London: Routledge.

Pettigrew, S. F. (2000) 'Ethnography and grounded theory: A happy marriage?', *Advances in Consumer Research,* 27, pp. 256-260.

Pink, S. and Morgan, J. (2013) 'Short-term ethnography: Intense Routes to Knowing', *Symbolic Interaction*, 36 (3), pp. 351-61.

Prior, L. (2008) 'Repositioning documents in social research', *Sociology,* 42 (5), pp. 821-836.

Prior, P. (1998) *Writing/disciplinarity: A sociohistoric account of literate activity in the academy*. Mahwah, NJ: Lawrence Erlbaum Associates.

Prior, P. (2005) 'Moving multimodality beyond the binaries: A response to Gunther Kress' 'Gains and losses.'', *Computers and Composition*, 22, pp. 23–30.

Pryor, J. and Torrance, H. (2000) 'Questioning the Three Bears: The Social Construction of Assessment in the Classroom', in Filer, A (ed.) *Assessment. Social Practice and Social Product*, London: Routledge: Falmer, pp. 110–128.

Rea-Dickins, P. (2001) 'Mirror, mirror on the wall: identifying processes of classroom assessment', *Language Testing,* 18 (4), pp. 429-462.

Robson, C. (*2002*) *Real World Research A Resource for Social Scientists and Practitioner-Researchers*. 2nd edn. Oxford Blackwell Publishers.

Rosmarin, A. (1985) *The power of Genre*. Minneapolis: University of Minnesota press.

Rowntree, D. (2006) *Assessing students: how shall we know them*? Abingdon: Routledge.

Sadler, D. R. (1989) 'Formative assessment and the design of instructional systems', *Instructional Science*, 18, pp. 119–144.

Salter-Dvorak, H. (2010) *Capturing the Ephemeral: An investigation of Second Language Masters Students' Quality of Participation in their Course Communities of Practice in Anglophone Academia*. PhD thesis. University of Bristol.

Salter-Dvorak, H. (2016a) 'Investigating Processes Underlying Identity Formation of Second Language Master' Students in UK Higher Educaton: insiders or outsiders?' in (Eds) Crossley, M. Arthur, L. and McNess, E. *Revisiting Insider-Outsider Research in Comparative and International Education*. Oxford: Symposium Books. pp 207-224.

Salter- Dvorak, H. (2016b) 'Learning to argue in EAP: Evaluating a curriculum innovation from the inside', *Journal of English for Academic Purposes*, 22, pp. 19–31.

Schmitt, D. (2015) 'Key Questions in EAP assessment', [Presentation], EAP Assessment: What's the score? Conference. St Andrews University, 28 February 2015. Available at: https://www.st-andrews.ac.uk/assets/university/english-language-teaching/documents/eap-conference/2015-presentations/diane-schmitt-key-questions-in-eap-assessment.pdf (Accessed Feb 2020).

Schmitt, D. and Hamp-Lyons, L. (2015) 'The need for EAP knowledge in assessment', *Journal of English for Academic Purposes,* 18, pp. 3-5.

Schmidt-Unterberger, B. (2018) 'The English-medium paradigm: a conceptualisation of English-medium teaching in Higher Education', *International Journal of Bilingual Education and Bilingualism*, 21 (5),pp. 527 -539.

Seidlhofer, B. (2005) 'English as a Lingua Franca', *ELT Journal*, 59, pp. 339-341.

Seidlhofer, B. (2011) *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Shah, A. (2017) 'Ethnography? Participant observation, a potentially revolutionary praxis'*, Journal of Ethnographic theory*, 7 (1), pp. 45-59.

Shohamy, E. (2001) *The power of tests: A critical perspective on the uses of language tests*. Harlow, England: Longman.

Shohamy, E. (2011) 'Assessing multilingual competencies: adopting construct valid assessment policies', *The Modern Languages journal*, 95 (3), pp. 418-429.

Shohamy, E. (2018) 'ELF and Critical language testing', in Jenkins, J., Baker, W., and Dewey, M. (eds), *The Routledge handbook of English as a Lingua Franca*. London: Routledge, pp. 583-593.

Slimani-Rolls, A. S., and Kiely, R. (2014). 'We are the change that we seek': developing teachers' understanding of their classroom practice'. *Innovations in Education and Teaching International*, *51* (4), pp. 425-435.

Spolsky, B. (1997) 'The ethics of gatekeeping tests: what have we learned in a 100 years?', *Language testing,* 14 (3), pp. 242-247.

Sundrarajun, C. (2007) *From interaction to presentation: Oral English skills development in the Thai university context.* PhD thesis. University of Bristol.

Sundrarajun, C. and Kiely, R. (2009) 'The oral presentation as a context for learning and assessment', *Innovation in Language Learning and Teaching*, 4 (2), pp. 101-117.

Swales, J. (1990) *Genre analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Swales, J. (1998) 'Textography: Toward a contextualisation of written academic discourse, *Research on Language and Social Interaction',* 31 (1), pp. 109-121.

Swales, J. (2019) 'The futures of EAP genre studies: A personal viewpoint', *Journal of English for Academic Purposes*, 38, pp. 75-82.

Taylor , L. (2014) 'Reflections on the "Issues Revisited" from the Perspective of a UK Examination Board', *Language Assessment Quarterly*, 11 (2), pp. 136-151.

Thurstone, L. L. (*1927*) 'A law of comparative judgment', *Psychological Review*, 34 (4), pp. 273–286.

Timonen, V., Foley, G. and Conlon, C. (2018) 'Challenges When Using Grounded Theory: A Pragmatic Introduction to Doing GT Research', *International Journal of Qualitative Methods, 17*, pp. 1–10.

Titscher, S., Wodak, R., Meyer, M., and Vetter, E. (2000) *Methods of text and discourse analysis*. London: Sage publications.

Torrance, H. and Pryor, J. (1998) *Investigating formative assessment: Teaching, learning and assessment in the classroom.* Buckingham: Open University Press.

Tsoldis, G. (2008) 'The (im)possibility of poststructuralist ethnography – researching identities in borrowed spaces', *Ethnography and Education*, 3 (3), pp. 271-281.

Tsushima, R. (2015) 'Methodological diversity in langauge assessment research: the role of mixed methods in classroom-based language assessment studies', *International journal of qualitative methods*, 14 (2), pp. 104-121.

Tummons, J. (2010) 'Institutional ethnography and actor–network theory: a framework for researching the assessment of trainee teachers', *Ethnography and Education,* 5 (3), pp. 345-35.

Tunstall, P. and Gipps, C. (1996) 'Teacher feedback to young children in formative assessment: A typology', *British Educational Research Journal*, 22 (4), pp. 389–404.

Van de Ven, A. *(2007) Engaged scholarship: A guide for organizational and social research*. New York: Oxford University Press.

Van Lier, L. (2004) *The ecology and semiotics of language learning: A sociocultural perspective.* Boston: Kluwer Academic.

Van Lier, L. 'The ecology of language learning: Practice to theory, theory to practice', *Social and Behavioural Sciences*, 3, pp. 2-6.

Vygotsky, L. S. (1978) *Mind in Society: The development of higher mental processes*. London: Harvard University Press.

Wajda, E. (2011) 'New perspectives in language assessment: The interpretivist revolution', in Pawlak, M. (ed.) *Extending the boundaries of research on second language learning and teaching* Heidelberg, Berlin, Germany: Springer, pp. 275–285.

Walford, G. (2008) *How to do educational ethnography*. London: Tufnell Press.

Walford, G. (2009) 'For ethnography', *Ethnography and Education,* 4 (3), pp. 271-282.

Wenger, E. (1998) *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.

Widdowson, H. (2020) *On the Subject of English: The linguistics of language use and learning.* Berlin: Mouton de Gruyter.

Wilson, R. (2016) *The PTE Academic and Outer Circle Students: Assessing Proficiency in English, Ownership of English, and Academic Performance at UK Universities*. PhD thesis. University of Warwick.

Wittgenstein, L. (1958) *Philosophical Investigations.* Oxford: Blackwell Publishers.

Wolf, D. P. (1993) 'Assessment as an episode of learning', in Bennett, R. E. and Ward, W. C. (eds), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment.* Hillsdale, NJ: Lawrence Erlbaum, pp. 213–240.

Wolf, D., Bixby, J., Glenn, J., and Gardner, H. (1991) 'To use their minds well: Investigating new forms of student assessment', in Grant, G. (ed) *Review of Research in Education,* 17, pp. *31-74.*

Wolfersberger, M. (2013) 'Refining the construct of classroom-based writing-from-reading assessment: The role of task representation', *Language assessment quarterly,* 10 (1), pp. 49-72.

Yates, J. and Orlikowski, W. (2007) 'The PowerPoint presentation and its corollaries: How genres shape communicative action in organizations', in Zachry M. and Thralls C. (eds), *Communicative practices in workplaces and the professions: Cultural perspectives on the regulation of discourse and organizations*. Amityville, NY: Baywood Publishing Company, pp. 67-92.

Zappa-Hollmann, S. (2007a) 'Becoming socialized into diverse academic communities through oral presentations', *Canadian Modern Language Review*, 63, pp. 455–485.

Zappa-Hollman, S. (2007b). *The academic literacy socialization of Mexican exchange students at a Canadian university*. PhD Thesis, University of British Columbia, Vancouver, Canada.

Zareva, A. (2009) 'Informational packaging, level of formality, and the use of circumstance adverbials in L1 and L2 student academic presentations', *Journal of English for Academic Purposes,* 8, pp. 55-68.

Zareva, A. (2011) 'L2 graduate student PowerPoint presentation designs: A reality check', *International Journal of Innovation and Learning,* 9 (2), pp. 127–144.

Zareva, A. (2016). 'Multi-word verbs in student academic presentations', *Journal of English for Academic Purposes*, 23, pp. 83-98.