Testing the Untestable or Defending the Indefensible: An investigation into EAP

assessment usefulness

By:

Julia Clifford

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

The University of Sheffield
School of Education

December 2016

ACKNOWLEDGEMENTS

ABSTRACT

With the Home Office permitting licenced Higher Education Institutions (HEIs) to decide how students' English language ability is assessed, many institutions use in-house English for Academic Purposes (EAP) assessments developed to reflect the particular context and focus of the HEI in which they are used. These are often created administered by and further developed by EAP practitioners. For this development to take place, it is necessary to establish the usefulness of such assessments to decide what development is required.

In an attempt to address gaps in the literature relating to research on in-house developed assessments, with a focus on how to establish whether these are useful to assess language proficiency, as well as providing students with the opportunity to develop wider academic skills, the research question for this study is: how can EAP practitioners establish the usefulness of in-use in-house EAP assessments? Bachman and Palmer's (1996) test usefulness framework is used as a basis for developing a practical framework for evaluating in-use assessments. The literature review considers EAP, language assessment and academic acculturation, ending with an evaluative framework for establishing authenticity of tasks, their predictive validity, and reliability of grading criteria and marking schemes.

The focus group case-study approach demonstrates how the framework can be given an institute-specific focus, considering opinions of three key stakeholder groups: EAP lecturers, subject-specialist lecturers and former pre-sessional students. Findings suggest academic language is highly contextual, and this context is institute- and, perhaps, department-specific. The findings suggest that for EAP assessments to be useful they need to support students with academic acculturation, providing valid and reliable results, supporting wider academic skills development.

Using the evaluative framework and establishing the institutional context in which students find themselves, this case-study of the EAP pre-sessional assessments used between 2010 and 2014 culminates in a transferable approach which can be used by practitioners before considering further assessment development.

CONTENTS

Chapter 4 – Methodology

PART II: FINDINGS AND ANALYSIS – IMPLEMENTING THE FRAMEWORK AND ESTABLISHING USEFULNESS

Chapter 5 – the PATs and the TLU situations

Chapter 6 – Predictive validity and reliability

Chapter 7 – Conclusions and Recommendations

Appendices

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EAP – English for Academic Purposes

HEI – Higher Education Institution

IEP – Intensive English Programme

MUC – Midlands University College

NNES – Non-Native English Speakers

NQF – National Qualifications Framework

PAT – Pedagogical Activities and Tasks

SELT – Secure English Language Test

SEP – Summer English Programme

TLU – Target Language Use

UKVI – United Kingdom Visas and Immigration

PART I: ESTABLISHING THE BACKGROUND AND DEVELOPING THE

FRAMEWORK

CHAPTER 1

INTRODUCTION

1.1 Some initial background

As a language learner and now also a language teacher and tester, language assessments have permeated my time in education; from my first spelling tests at primary school to the English for academic purposes (EAP) assessments I now find myself devising, revising and marking, there seems to be no escape. Yet I do not consider this inability to escape to be a bad thing. I have always enjoyed the sense of achievement that language assessments have given me, and now I also enjoy assessing students to show them their own progress, and also as a way of helping them to develop skills and abilities which they will need in the situations outside the classroom. So, since there is clearly no escape from language assessments, what I have found to be necessary is a way of establishing whether or not the assessments are actually testing what they need to; basically, I want to establish a way of deciding if a test already in use is useful.

But useful to whom? There are many reasons why language tests are used, and Douglas (2010, p. 17) explains that these uses include (but are not limited to):

> to determine whether learners have mastered the material presented in class, to discover areas of weakness in learners' language knowledge so that their needs can be addressed, to place learners in appropriate instructional levels, or to find out whether learners know enough of a language to use it for some purpose outside the language classroom.

The focus of this thesis is English for academic purposes (EAP), and the purpose of the assessments used on EAP programmes is, as will be discussed in Chapter 2, complex. But in terms of to whom they should be useful, then they certainly need to be useful to the students, since this is the easiest way for them to see their own progress. However, assessments also need to be useful to other stakeholders: the teachers who administer

them and provide feedback and progression recommendations on the basis of results; and the lecturers who will be teaching the students when they progress onto their main programmes of study. Already the multifaceted nature of the EAP assessment is becoming clear, and a more detailed definition of EAP and its purposes will be presented in Chapter 2.

So, what will this study be offering? Well, before I come to that, I feel it necessary to explain what it will not be offering. Firstly, the focus of this study is not to evaluate the teaching methods being used on the courses, since this has been done by numerous authors over the years (see, for example Ypsilandis and Kantaridou, 2007; Watson Todd, 2003; Read, 2002; Cumming *et al.*, 2000; Enright *et al.*, 2000; and Jordan, 1997). Further, this study has not been designed to evaluate the course content since this area has also been researched extensively in relation to EAP (for example Donohue and Erling, 2012; Evans and Morrison, 2011; Guse, 2011; Dooey, 2010; Spector-Cohen *et al.*, 2001; Ferris and Tagg, 1996a; and Ostler, 1980). Additionally, this is not an action research project since the starting point of this study is not a problem which needs research to solve it, but rather it is the step which comes before this; it is a way of establishing whether there actually is a problem which needs solving.

One of the issues faced by EAP practitioners is predicting the future academic and linguistic needs of students in order to design courses which will help prepare these students for life both in and out of the classroom. These predictions are also used to create assessments which should usefully test whether the students have acquired the skills identified to a sufficient level to be able to cope on their main programmes of study (Ypsilandis and Kantaridou, 2007). Therefore, the focus of this study is to establish a framework which can be used to evaluate the assessments used on EAP pre-

sessional programmes – the in-house[1] assessments – to investigate whether they are useful. In order to do this, as well as creating an evaluative framework, which will emerge from the literature review, it was also necessary to consult three key stakeholder groups: the EAP staff, the pre-sessional students and the teaching staff on the receiving programmes. By using their opinions to extend and explain the framework and combining these with definitions provided by the key literature, it was possible to evaluate the assessments written by members of staff at the institution and used on programmes between 2010 and 2014 (henceforth in-use assessments) to establish just how useful they were to those who they have the biggest impacts on: the English teaching staff, the students and the subject-specialist lecturers on the main programmes of study onto which students can enrol once they complete the English programmes.

1.2 Research question and objectives

Now I have briefly introduced the study, I come to the overarching research question for this thesis: how can EAP practitioners establish the usefulness of in-use in-house EAP assessments? To help answer this question, the following objectives have been set:

- To provide a brief history of English for academic purposes

- To critically evaluate the literature relating to language assessments, assessment evaluation and academic acculturation

- To develop an evaluative framework to establish the usefulness of EAP assessments

- To investigate EAP staff, student and subject specialist staff opinions regarding the language and wider skills required for coping with HE level study

---

[1] These in-house assessments are those which are created and developed within the institution for use on programmes at that particular institution, rather than the assessments developed by examination boards such as IELTS and Trinity.

- To carry out an audit of subject-specific assessments in use at the case-study HEI to establish typical assessment tasks in use between 2010 and 2014

- To use the opinions of EAP lecturers, subject specialist lecturers and former EAP students along with the results of the assessment audit to define and expand the key terms in the evaluative framework

- By amalgamating the findings from the primary and secondary research, to critically evaluate the EAP assessments being used at the case study HEI between 2010 and 2014

Before I start to work on my objectives, however, I feel it necessary to partially unpack the main research question, starting with an overview of the Home Office regulations which relate to supporting visa applications to overseas students. 2016 Home Office regulations state that any student who requires a Tier 4 visa to study in the UK needs to provide proof of their English language ability. For those students who wish to study a programme which is below level 6 of the national qualifications framework (NQF), for example a foundation degree, the institution must use one of the Secure English Language Tests (SELTs) from an approved provider and taken at an approved test centre, that is assessments from IELTS or Trinity (Home Office, 2016). However, for those students who wish to study a programme leading to a qualification at NQF level 6 or above,

> if your sponsor is an HEI, we will allow them to choose how to assess your English language ability. … Your sponsor must … ensure that you are proficient to level B2 on the Common European Framework of Reference for languages (CEFR) in each of the four components (speaking, listening, reading and writing).
>
> Home Office (2016, p. 37)

In the case study HEI, for students on pre-sessional English programmes and usually progressing onto courses at NQF levels 6 or 7, this level of English ability is assessed

using EAP assessments devised, revised and administered by the staff at the HEI. Those being discussed here were used at the end of the in-sessional and pre-sessional English language programmes between 2010 and 2014. These are what Davies (1977) terms achievement assessments and they test what has been learnt of a known syllabus. This differentiates them from proficiency tests which assess "what has been learned of a known or unknown syllabus" (Davies, 1977, p. 46). The distinction between the two, however, is not quite a clear-cut as Davies' definitions would have us believe, though, so I will discuss these differences in more detail in Chapter 2.

In addition, it is necessary to establish why I am focussing on EAP practitioners rather than on language assessment experts. Basically, it is these EAP practitioners, the teachers who work with, teach and support the students throughout their time on their pre-sessional programmes, who often create, devise and grade the assessments. They are not usually testing experts, nor do they necessarily have the time, resources or expertise to carry out statistical validation and reliability studies on the tests they devise and grade. They are specialist teachers, that cannot be ignored; but they are not statisticians, nor are they expert test writers. They do, however, have an in-depth understanding of the needs of international students coming to study in the UK HE environment, and they are aware of the linguistic demands of HE study. They are also, as the title of this thesis implies, often called on to defend the grades awarded on the basis of their assessments; grades which, to some, appear indefensible. Additionally, they are expected, as will become clear throughout the discussion, to test not 'just' the language, but myriad other skills; testing skills which seem at times untestable.

Finally, I feel it necessary to explain why I have chosen the term 'in use'. In the majority of institutions where in-house assessments are used, it is rare that EAP

practitioners will develop whole new suites of assessments from scratch; instead personal experience has shown that they will be revising and improving existing assessments. Thus a framework which focusses on how to establish whether assessment are useful prior to deciding if, how and why to further develop them is something which could be invaluable in terms of in-house assessment development. There are many models of assessment validation, reliability tests and even usefulness frameworks which already exist, as will become clear in the literature review, however these for the most part relate to the development of new assessments. There is very little which has been developed for the revision of in-use assessments, hence the current study.

1.3 Rationale

Having explored the research question in more detail, it is now necessary to further explain the relevance of this research project. As explained by Clovis (2007, p. 137) "assessment is one of the most difficult, time-consuming, and complex parts of teaching". In fact, Hatipoğlu (2010) highlights that "the typical teacher can spend as much as a third to a half of his or her time involved in assessment-related activities" (p. 40). However, as Norton (2007, p. 93) explains

> assessment is sometimes the last thing that we think about when designing our courses. We tend to think about curriculum and what should be covered and only when that has been determined do we turn our attention to how we might assess what our students have learnt.

Taking this as a starting point, then, there were three reasons which made this research necessary. Firstly, as inferred in the previous section, the importance of English language testing cannot be underestimated since it has a significant impact on both learners and teachers. For learners, their English language grades are often a deciding factor of whether they can study at their chosen overseas university; for English teachers, the assessments partially drive what is taught on the English language programmes; for the subject-specialist lecturers, the grades derived from the English

language assessments provide them with an indication of their international students' abilities, both in terms of their language proficiency and their wider academic skills. In fact McNamara (2000) explains that having "an understanding of language testing is relevant both for those actually creating language tests, and also more generally for those involved in using tests or the information they provide" (p. 5). Thus, investigating the usefulness of the assessments in use at the case study HEI will impact on all three key groups of test-users. Therefore, in order to establish the usefulness of the tests for these three groups, a sample from each group was invited to participate in the research[2].

Secondly, from a more personal point of view, although often in contact with language assessments, my interest in this topic was really first piqued by a change in office where, for the best part of three years, I shared an office with the HE assessment team at the case study institution. Sharing this office meant that I was made more aware of assessment practices within the institution. As a result, I began to pay closer attention to the assessments and tests being written for and administered to students on the English programmes. This increased awareness led to the thesis which is presented here, since it sparked in me a desire to establish just how useful our assessments are for the main test users. My interest was further roused by a small workshop which the HE assessment managers asked me to run during the annual institution-wide training days in July 2012 and which focussed on what IELTS scores meant in practice for our students. The discussions which took place during the workshop resulted in me wanting to further investigate the English language assessments in use at my place of work, with a view to establishing if there was a match between the skills they were assessing and developing the skills deemed necessary for students to be able to cope on their HE courses. Having briefly looked at these, I quickly realised that a more in-depth

---

[2] For further discussion of this, see the Methodology in Chapter 4

approach was required, and that an evaluative framework would be needed if an informative and helpful evaluation of their usefulness was to be carried out; hence the current project.

Further, as Ostler (1980, p. 489) explains:

> academic populations change over time, suggesting periodic assessments of the current students' requirements and abilities to see if changes in the population have also brought about a change in needs.

Despite annual reviews of assessments in the form of module evaluations and the institutional module review process, in my 14 years at the HEI, an in-depth evaluation of the assessments, and more importantly of their usefulness, had not been carried out. This study addresses this need since, to paraphrase Ostler (1980) further, it is necessary to establish whether or not the EAP assessments in use are serving the real needs of both students and staff.

Finally, there is the question of where this study sits in relation to the literature in the field. Schmitt and Hamp-Lyons (2015) recently edited a special edition of the *Journal of English for Academic Purposes*. In their editorial, they explain that they were hoping that practitioners would write about "local, in-house assessment" (p. 3) since "large-scale, standardised tests … have attracted the bulk of the research in EAP assessment" (*ibid*.). They also highlight that programmes such as those considered in this study "aim to orient students to the academic conventions and culture of the local HE system" (Schmitt and Hamp-Lyons, 2015, p. 4). Hence, a study which considers the usefulness of these assessments in doing this will offer a useful contribution to the field.

Essentially, this study addresses a real need in the field of EAP assessment, in particular in relation to the evaluation of EAP assessment usefulness; something which is often investigated, but which seems to be carried out, at least at the case study HEI, in an ad-

hoc, non-systematic fashion, and as part of the compulsory annual review. Informal discussions with other EAP lecturers suggests that this is also the case at other institutions too, and so a practical framework for evaluating the usefulness of in-use in-house EAP assessments is something which could be of value to various institutions.

1.4 Introducing the case study HEI

In order to design useful tasks to help prepare college or university students for their mainstream courses, it is necessary to consider "the type of institution in which they are or will be studying" (Ferris and Tagg, 1996a, p. 47), thus in order to further contextualise this study, this subsection will provide some key details about the case-study HEI. The HEI[3] is a vocationally-focussed university in the Midlands of England which specialises in teaching hospitality, tourism, culinary arts, hair and beauty, education and sports therapy from foundation degree to masters' level. It also offers FE courses in the same areas. MUC was granted full university status in 2012, and its reputation overseas is such that a large number of the students on the courses are international, with the vast majority coming from East Asia, more specifically China and Hong Kong. There are also increasing numbers of students coming from the Indian sub-continent and Saudi Arabia, and in total the University welcomes over 1,100 students from 65 different countries from around the world. For an institution which has a total of just 8,000 students across all courses from FE to postgraduate, this number of international students is not insignificant. Being able to usefully assess these students' abilities in the skills they need to cope at HE level is, thus, important, not just for the reputation of the University, but for the benefit of the international students they work with each year.

---

[3] From here on to be known as Midlands University College or MUC

The English language department of the University is very small considering the number of students which it deals with each year. At the time of study, there were five full-time English language lecturers, who were supported by sixteen agency staff working on various short-term contracts throughout the year. The University offers pre-sessional English language programmes throughout the year, as well as in-sessional courses, subject-specific support and the English modules on an international foundation programme. In terms of the pre-sessional programmes, these run during the main teaching year, with students able to study for 13 or 26 weeks depending on their English level on arrival. Between 2010 and 2014, the University welcomed approximately 30 students per year onto these programmes. In addition to this, the main pre-sessional programmes for the university run during the summer, with students able to study a 10-week or 6-week English language pre-sessional between the end of June and the start of September. During the time of this research study, some 400 students were enrolled on the summer pre-sessional programmes. All pre-sessional students were assessed, as is the institutional norm, at the end of their programmes using assessments developed in-house.

The aim of the pre-sessional English programmes is to provide students with the opportunity to improve their English proficiency level prior to starting their main programmes of study. The entry requirements in terms of English language for the pre-sessional programmes are presented in the table below:

Table 1 – Pre-sessional English Programme Requirements at MUC

Programme length	26 weeks	13 weeks	10 weeks	6 weeks
English entry requirement	IELTS 4.5 with no less than 4.0 in any skill	IELTS 5.5 with no less than 5.0 in any skill	IELTS 5.0 with no less than 4.5 in any skill	IELTS 5.5 with no less than 5.0 in any skill

Each of the pre-sessional programmes has the same learning outcomes:

- To provide English language training skills and practice, and to provide a sound basis for future studies at [MUC]

- To equip students with the necessary academic skills to study up to postgraduate level in a UK higher education establishment

- To provide experience of being taught English language in a UK higher education establishment

- To encourage participation in a diverse range of teaching and learning experiences

- To encourage students to use communication skills in a higher education setting and in social situations

- To enable students to develop knowledge and understanding of approaches to study / culture in a UK higher education establishment

As can be seen from these learning outcomes, the programmes are attempting to help students their language skills, as well as the academic skills to cope with HE level study in the UK and to adapt to life at their new institution.

In terms of my role within the institution, I have worked at MUC since 2003, mostly in the English Development Centre, but also as a member of the modern foreign languages team as well as the masters dissertation team. In the English team, I am an experienced and long-standing member of staff at MUC, however I am considerably less experienced than other members of the teaching team in terms of English language teaching. Since starting at MUC, I have worked on all areas of assessment: using,

marking, writing and developing. I am, therefore, extremely familiar with the assessments being investigated in this study.

1.5 Thesis structure

This study has been split into two parts: Part I offers the theoretical background and methodological approaches leading to the development of the evaluative framework; Part II demonstrates how the framework can be applied in practice. The supporting objectives presented above offer a general overview of the structure of the thesis, and the next chapter, Chapter 2, will provide a literature review, to be followed in Chapter 3 by a discussion of theories and models leading to the evaluative framework which is proposed for the evaluation of in-house in-use EAP assessments. Following this, Chapter 4 will discuss how I collected the stakeholder opinions, as well as outlining how I gathered institution-specific information pertinent to the evaluation of the assessments. Part II of the study starts with Chapter 5, in which I use the stakeholder opinions, along with the key literature, to explain the assessment usefulness in terms of academic acculturation. The second discussion chapter, Chapter 6, will offer an evaluation of the assessments in use at MUC between 2010 and 2014 in terms of their validity and reliability. This will be followed by conclusions and recommendations in Chapter 7.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Although language testing "has a long history as a social practice" (McNamara and Roever, 2006, p.1), it has "only a relatively short one as a theoretically founded and self-reflexive institutional practice" (*ibid.*). It does, however, have a great deal of literature relating to it, and it is on this literature that I will be drawing to construct this section of the thesis, with the hope of adding to the aforementioned relatively short theoretically founded and self-reflexive institutional practice. To do this, this section will begin with a brief discussion of the origins of EAP; following this, I will offer definitions of key terms relating to the evaluation of language ability and an outline of the history of language assessment so as to offer some context for the study. The ideas will be drawn from three key areas: language assessment and testing, assessment in HE and literature relating to EAP. Following this, I will consider literature relating to academic acculturation, which is a key part of EAP teaching, as well as considering the more general area of academic literacies. The review will be organised as answers to a series of questions relating to the general area of language assessments: What is EAP? What is the link between EAP and academic acculturation? Is there a difference between home and international students? What is assessment in HE? What is a language assessment? and finally, Why use language assessments?

2.2 What is EAP?

EAP started life in the 1970s as one of two main branches of English for Specific Purposes (Benesch, 2001), the other being English for Occupational/Vocational/ Professional Purposes (EOP/EVP/EPP) (Ypsilandis and Kantaridou, 2007). EAP, in

fact, emerged from a "shift towards a strong focus on *language for communication …* particularly in the context of higher education and vocational training as growing numbers of international students sought access to educational opportunities in English-medium institutions" (Taylor and Geranpayeh, 2011, p. 90, original emphasis). As an academic discipline, then, EAP is relatively new, and the literature suggests that it has been in existence for some forty years (Hyland and Hamp-Lyons, 2002). In that time, however, there have been numerous shifts in approaches to researching and writing about it (Taylor and Geranpayeh, 2011). Since its early days as a separate strand of English language teaching there have been a great deal of changes in EAP teaching practices, teaching materials, the use of technology, assessment practices and teacher-training (Hyland and Hamp-Lyons, 2002; Jordan, 2002; Watson Todd, 2003).

There are several reasons for the growth of EAP over the past forty years, but key among them seems to be that English has become "the leading language for the dissemination of academic knowledge" (Hyland and Hamp-Lyons, 2002, p. 1). This coupled with the fact that "English has become an even more widespread international language, for business and higher education" (Jordan, 2002, p. 69) has meant that English is now taught all over the world by both native and non-native speakers for and in various contexts. As a way to differentiate EAP from more general English programmes, Hamp-Lyons (2001, p. 126) explains that "[EAP] begins with the learner and the situation, whereas general English begins with the language". Hyland and Hamp-Lyons (2002, p. 3) add that EAP

> draws on a range of interdisciplinary influences for its research methods, theories and practices. It seeks to provide insights into the structures and meanings of academic texts, into the demands placed by academic contexts on communicative behaviours, and into the pedagogic practices by which these behaviours can be developed [and assessed]

The term EAP was, in fact, first used in the mid-1970s, and was in general use within the British Council by the spring of 1975 (Jordan, 2002). It was in the 1980s that the English Language Testing Service (ELTS) first introduced a "task-based, discipline-specific and integrated approach to testing language skills needed for university study" (Taylor and Geranpayeh, 2011, p. 90), which they also termed EAP. More recently, Tajino *et al.* (2005) used a soft-systems methodology to provide a definition of EAP, which they explain is:

> a productive combination of English for general education context (e.g. other cultures and international issues) and practical English language skill-building, which can meet the requirements of concerned people, and which can be implemented with the resources available in the EAP course.
>
> Tajino *et al.* (2005, p. 35).

Watson Todd (2003, p. 153) in explaining how EAP differs from general English states that "perhaps the most obvious differentiating characteristic of EAP is the needs of the students and thus the content and goals of teaching" adding that

> in EAP, these needs, and thus the teaching [and assessment] relate to a study purpose … and EAP is largely founded on the fact that the English used to fulfil the study needs stand in contrast to general English.
>
> Watson Todd (2003, p. 153).

Taylor and Geranpayeh (2011) also touch on the idea of needs analysis, explaining that in assessing EAP, what is required is a test which is "based on a careful 'needs analysis' of target language use in context" (p. 90)

In looking at EAP teaching, research by various authors demonstrates that EAP courses provide instruction on more than just English language (Watson Todd, 2003; Hyland and Hamp-Lyons, 2002; Jordan, 2002; and Benesch, 2001). Watson Todd (2003) explains that "the main goal of EAP is for students to communicate effectively in academic environments" (p. 149), and a key factor for achieving this goal is "knowing what the communicative requirements in these environments are" (*ibid.*), as well as

knowing how best to assess students to ensure they will be able to cope with these communicative requirements. Tajino *et al.* (2005) add that EAP courses teach a combination of language skills whilst also offering students a way of learning about other cultures and helping them to broaden their understanding of international issues.

This brings me to the final point for discussion in this subsection: Why is the EAP aspect so important to this study? In answer to this question, it is necessary to comment on Bachman's (1990) assertion that it is vital to consider "context beyond the sentence" (p. 4), context which "includes the discourse of which individual sentences are part and the sociolinguistic situation which governs, to a large extent, the nature of that discourse in both form and function" (*ibid.*). What this means, for Bachman at least, is that language tests need to assess communicative language ability, and should:

> not only measure a wide range of language abilities, including grammatical, discourse, sociolinguistic, and strategic competences, but that are also 'authentic', in that they require test takers to interact with and process both the explicit linguistic information and the implicit illocutionary or functional meaning of the test material
>
> Bachman (1990, p. 4)

Following on from Bachman, Davies (1990, p. 51) defines language thus: "like all human behaviour, [it] is inexact", adding that "all measurement, even the most rigorous scientific measurement, lacks absolute precision" (*ibid*.), yet it is demanded. He also explains that:

> there are two main senses of language ability. The first is the more general sense in which we make distinctions in terms of knowledge or performance [...] The second sense is more often used in relation to second- or foreign-language learners: here we take account of some level of native-speaker ability to which we relate the performance of second-language learners.
>
> Davies (1990, p. 52).

Finally, Bachman and Palmer (1996, pp. 61 – 62) explain that:

> in general, *language use* can be defined as the creation or interpretation of intended meanings in discourse by an individual, or as the dynamic and interactive

negotiation of intended meanings between two or more individuals in a particular situation.

I do, however, need to partially adapt this definition as it is about language use *in general* rather than EAP, but it is EAP that is the focus of my study. For this study, then, what is needed is a definition of academic language use, which I define thus: the creation or interpretation of intended meanings in *academic* discourses by an individual, or as the dynamic and interactive negotiation of intended meanings between two or more individuals in various *academic* situations. My adapted definition reflects that of Bailey and Huang (2011) who, although defining English for school use rather than university use still provide a helpful definition of academic English. They explain that academic English is "the oral and written language used in classrooms by teachers and students for the purpose of classroom management and academic learning, as well as the language of textbooks, assessments, and other curricular materials" (Bailey and Huang, 2011, p. 349). This distinction between general and academic English is particularly important since, as Hyland and Hamp-Lyons (2002, p. 2) explain:

> teaching [and assessing] those who are using English for their studies differs from teaching those who are learning English for general purposes only. It is also different from teaching those who are learning for occupational purposes.

2.3 What is the link between EAP and academic acculturation?

As the above definitions of EAP highlight, EAP focuses on the academic situation in which the incoming students will find themselves. A great deal has been written about the differences between educational systems, and this section will focus on the link between EAP and academic acculturation.

Academic acculturation is defined in various ways by different authors in the field. In reviewing literature in the field of acculturation, Smith and Khawaja (2010) explain that originally it was conceived as the process migrants go through of "discarding their

cultural heritage and adopting the host country's culture" (p. 701). Bang and Montgomery (2013) also provide general definitions of acculturation, explaining that it "refers to an individual's adjustment process and constitutes the outcome of the interactional dynamics between two or more cultural systems" (p. 344). In terms of international students in their new academic environment, these definitions need some slight adaptation. Thus, academic acculturation can be seen as an international students' adjustment process which constitutes discarding the academic cultural heritage of their home country and adopting the host country's academic culture.

This new academic culture can be seen in approaches to teaching, approaches to assessment, classroom behaviour, group work requirements or even things as simple as how to address a lecturer (Cheng and Erben, 2012; Dooey, 2010; Smith and Khawaja, 2010; Wong and Thang, 2008; Andrade, 2006; Watson Todd, 2003; Jordan, 1997). It is also seen in the large differences between overall teaching philosophy of the country (Ramachandran, 2011; Kingston and Forland, 2008; Kember, 2000). How all of this links to EAP is explained by Dooey (2010) who comments that one student involved in her study "seemed to view the [English language centre] as a sort of training ground for future challenges" (p. 192). This, and much previous and subsequent research implies that the pre-sessional English programmes which students undertake prior to starting their main programmes of study are where the process of academic acculturation begins.

In offering a definition of acculturation, Cheng and Erben (2012) explain that it is "an individual's process of adaptation and socialization to the host culture" (p. 479), adding that "during the process of acculturation, recent arrivals learn the language and values of the target culture" (*ibid*). In terms of EAP programmes and assessments, Dooey (2010, p. 186) comments that:

international students … often suffer from a range of problems related to their adjustment to a new academic and cultural environment. Therefore the role of preparation courses [and their assessments] is crucial, not just in terms of developing English language skills at a macroskill level, but also in fostering among students the vital … skills required for successful intercultural communication and integration into university life.

Cheng and Erben (2012), in fact, explain that when international students begin a course overseas, "not only [do] they enter a new program of studies, but also they connect … with a distinctly different culture simultaneously" (p. 490). Andrade (2006) also highlights the need for students to become familiar with institutional norms, rules and regulations, since as Ramachandran (2011) explains some international students "find themselves unable to understand and address the demands of a different system" (p. 207). Ramachandran (2011) also explains that many international students are often "baffled by the jargon used by their peers, teachers and academic administrators" (p. 207), adding that "international students need to be familiar with the common jargon and expressions used in the British [HE] environment" (p. 213).

A further area of academic acculturation is the new approach to teaching, which for many international students can be quite a shock. Ramachandran (2011), for example, explains that unfamiliar pedagogical approaches include "…group activity, presentation and other student-led initiatives" (p. 204). Cheng and Erben (2012) also comment on this, explaining that differences in approaches to teaching and assessment include the fact that students need to "participat[e] in group discussions and projects, mak[e] oral presentation and discuss… topics in seminar formats" (p. 491). Further, Andrade (2010) highlights that for students in her study "classroom discourse patterns present challenges as these differ culturally, and international students may be uncomfortable with what they view as [home] students' lack of respect for the professor" (p. 223). Thus, there is a significant difference in relation to how classes are taught and the

amount of participation expected from students in UK HEIs such as MUC when compared with students' home institutions and previous educational experience.

Related to the different approaches to teaching are the potentially very different approaches to learning expected of students in Western universities, and Dooey (2010, p. 193) emphasises that pre-sessional English programmes are:

> a valuable provider of language-related academic skills (for example, acknowledging sources, referencing conventions, academic style, and register) over and above the technical aspects of the language.

Ramachandran (2011), in looking in detail at UK higher education comments that:

> …HEIs emphasize that students must participate in and contribute to the learning process. With text-based and memory-based studies predominant in several countries and academic systems, often international students have no experience of an effort-based system… they are more familiar with teacher-led paradigms
> <div align="right">Ramachandran (2011, p. 207)</div>

This is reflected by Lilyman and Bennett (2014) who explain that "…international students may find that they need to develop different learning strategies and study patterns from those used in their own countries" (p. 69). Further, Wong and Thang (2008) highlight key skills for academic achievement in HE as "classroom based oral presentations skills, understanding lectures, note-taking and note-taking skills, academic reading and vocabulary, and academic speech and pronunciation" (p. 5).

In terms of the differences between Eastern and Western educational traditions, Kingston and Forland (2008) highlight the potentially large discrepancy between "the collectivist educational traditions found in East Asia (based on Confucian philosophy) and the individualist educational tradition found within Western societies (based on Socratic philosophy)" (p. 204). These large differences can lead to international students, particularly those from East Asia, experiencing difficulties in their transition to

the new educational culture especially "with regard to underdeveloped skills for autonomous study, research skills (particularly with regard to utilising the library), language competencies and Western lifestyles" (Kingston and Forland, 2008, p. 210). The pre-sessional programmes, and their corresponding assessments, then, need to provide students with the opportunity to acquire, develop and practice these key academic skills, and to demonstrate that they have mastered them sufficiently so as to be able to cope with their future studies.

Another key area of difference in terms of learning is that of autonomy, something which may be new to some international students. Yet, as Skyrme (2010) explains in Western higher education, "the university value[s]… the autonomous student who can solve academic problems without undue reliance on teachers" (p. 215). Wong and Thang (2008) also consider autonomy, commenting that among the purposes of EAP programmes should be "helping students acquire appropriate learning strategies to promote learner autonomy" (p. 3). Watson Todd (2003) touches on this point too, calling for EAP to promote autonomy in international students. Additionally, Jordan (2002) comments that the expectations of students coming to study in the UK included "expecting to be told precisely what to read by their subject tutors, who they also expected to correct any mistakes they made with English" (p. 75).

Autonomy, then, is a key aspect of UK HE which incoming international students may need support in developing. Students from certain educational backgrounds are not used to the more autonomous approach, and in one of the studies reviewed by Andrade (2006), it was found that "[international] students did not take responsibility for their own learning" (p. 138). She further explains that tutors in her study criticised students,

stating that they did not "take responsibility for their own learning" (Andrade, 2006, p. 138), whereas the students "found professors indifferent" (*ibid.*).

In terms of learning, an additional area for discussion is the critical element which is required by most Western institutions. Smith and Khawaja (2010) explain that students in their study had difficulties adjusting to "the interactive teaching style and critical approach to learning" (p. 703) in Western-style HEIs, adding that "students from countries which focus upon rote learning may find it particularly difficult to adjust to the importance placed on critical thinking" (*ibid.*). This critical approach to learning is also discussed in the myriad study skills handbooks available for students entering UK HE (see for example, Cottrell, 2013; McMillan and Weyes, 2012; McMurray, 2011, Moore *et al.*, 2010; and Lewis and Reinders, 2003), thus emphasising its importance for coping with HE level study.

The final area of acculturation is assessment acculturation, or literacy. In terms of the new approaches to assessment, Cheng and Erben (2012) highlight that a significant difference between the Western and Eastern educational assessment culture is that "grades [are] based on … performances in a variety of class activities rather than simply on a few exams" (p. 491). In fact, as Phakiti *et al.* (2013) explain "many international students … encounter difficulty in grasping the specific purpose of academic assessment tasks" (p. 243). This could be due to the rote learning and memorisation which feature highly in many Eastern educational traditions (Kingston and Forland, 2008; Kember, 2000). This significant difference in assessment approaches can lead to confusion and discomfort when students are expected to apply rather than reproduce knowledge.

Focussing on written assessments in particular, Lebcir *et al.* (2008) comment that there may be "specific gaps in expectations between international students and academic tutors" (p. 270). These differences are further explained by Bacha (2010) who comments that many international students "have difficulty producing academic arguments in their required essays" (p. 229). In explaining what academic argument is, Bacha outlines it as "includ[ing] a writer's opinion (or claim) and support (evidence)" (p. 230), an approach to writing with which many international students are unfamiliar. This is also commented on by Zhang and Mi (2010) who explain that:

> there are prominent differences between cultures in terms of rhetorical conventions (ways of presenting ideas and developing arguments), cultural schemata (whether one transforms/extends knowledge), and writing perspectives or expectations (whether it is the writer's responsibility or the reader's responsibility to make sense of the text).
>
> Zhang and Mi (2010, p. 385)

Additionally, to extend the comment from Bacha (2010), Zhang and Mi (2010) highlight that "what constitutes evidence for an argument, and what argument is are … culturally determined" (p. 238). Moreover, Andrade (2006) highlights in a review of studies about international student adaptation that many students are "unaccustomed to analysing the strengths and weaknesses of an argument" (p. 138). How to construct a written assignment effectively, then, is another important element of coping with UK HE.

Thus far, the discussion has presented rather stereotypical views of students, both home and international. The stereotypes can be summed up into three key areas, as highlighted by Kumaravadivelu (2003): international students "(a) are obedient to authority, (b) lack critical thinking skills and (c) do not participate in classroom interaction" (p. 710). This is supported by Ha (2006) who explains that typical stereotypes expressed by Western academics are the international students are often

obedient to authority and lack critical thinking skills. Hammersley-Fletcher and Hanley (2016) also take issue with this type of stereotyping, commenting that "international students are often stereotyped as 'rote learners' in a Western environment where techniques such as memorisation and respecting the voices of experienced educators are not valued" (p. 979). They also comment that this often leads to "a deficit situation, where South-East Asian students do not engage in deep learning" (*op. cit.,* p. 983).

In addition to the issues of stereotyping noted above, as with many researchers in the area of academic acculturation, I have used the terms East and West as a homogenous dichotomy approach to discussing educational norms and cultures. Kumaravadivelu (2003) calls this approach "a harmful homogenisation of nearly 3 billion people" (p. 710). I have not, however, made this decision uncritically, but this is an area which requires a little more explanation. On the advice of Durkin (2008) I have made this choice "with caution recognizing the complexity of the issues" (p. 16). I am, for example, aware of the differences in teaching and learning styles across Asia, and am aware that simply using the labels 'Eastern' or even 'Confucian heritage cultures' is very simplistic. Further, I am aware that teaching styles across Asia vary considerably (Kumaravadivelu, 2003). I am also aware that even in those countries which are often labelled 'Confucian heritage cultures', that is those countries which make up the eastern region of Asia, also have different approaches to teaching and learning (Kumaravadivelu, 2003; Ryan and Louie, 2007; Durkin, 2008; Sigurðsson, 2017). In the same way, I am also conscious that 'Western' styles of education vary considerably, with each different country having its own requirements regarding, for example, referring to the work of others in writing. That being said, however, Durkin (2008) does explain that "a sufficiently identifiable core of rhetorical traditions which will allow for the use of the singular label 'East Asia' does seem to exist" (p. 16). This is

further supported by Mason (2007) who explains that, in terms of the East/West dichotomy, education in these different regions "is informed by widely different historical and cultural perspectives, from western to Confucian, from liberal to communitarian, from colonial to postcolonial" (p. 339). Although broad generalisations, then, I have chosen to use the terms East and West when discussing how EAP can aid with academic acculturation.

All of the discussion of international students so far however, as well as providing possibly unhelpful and inaccurate stereotypes and over generalisations of two broad educational traditions, seems to take a 'deficit' view of these students (Lea, 2004; Goldingay *et al.*¸2016). This view sees them "in terms of the characteristics that they lack, rather than those which they bring to their new environments" (Ryan and Louie, 2007, p. 406). Many authors, in fact, discuss East Asian students taking an almost polar opposite approach. Kumaravadivelu (2003), for example, explains that it is simply wrong to assume that "Asian students do not think critically and that certain values underpinning the notion of critical thinking are incompatible with their cultural beliefs" (p. 712). Certainly, I, like Kumaravadivelu (2003), do not believe that Western students are naturally good at critical thinking because "there is something inherent in the Western culture that promotes critical thinking" (p. 712), however, the need to help students begin to develop a Western take on critical thinking is something which is valid for inclusion in a suite of EAP assessments.

A particular proponent of this opposite view is Sigurðsson (2017), who highlights that "Confucianism emphasises the importance of a continuously critical attitude from which we, products of contemporary Western culture, may have much to learn" (p. 132). Although a useful source for refuting the 'myth' that East Asian educational

traditions do not encourage criticality, it is more concerned with comparing early Confucian philosophy with current Western approaches; however this is not necessarily a helpful comparison to make. This opposite polarisation is discussed by Ryan and Louie (2007) as being as unhelpful as the deficit view of students, and it is to Chanock (2010) who I turn for greater explication of the issues. She explains that the stereotype that East Asian students "are uncritical: that they accept what they told and 'parrot it back'" (p. 547) is inaccurate adding that these students are regularly critical, but that this is rarely explicit in their writing. This is not necessarily a matter of limited critical ability, rather it is a "matter of manners" (Chanock, 2010, p. 547); essentially, in order to save the face of the person they are disagreeing with, they will often present the disagreements in a form which is "so muted as to seem like an aside" (Chanock, 2010, p. 547). In terms of how EAP can aid students in their academic acculturation, then, it is important that the assessments used help students to develop and demonstrate these critical skills, with students required to carry out tasks which mirror those with which they will be faced when they progress.

But how does all of this link to EAP? Wong and Thang (2008) explain this when they say that in EAP "students are exposed to the expectations and requirements of their [new institution]" (p. 4). Andrade (2006) also comments on this, noting that "students' writing was inextricably bound by their cultures, ways of seeing the world and identities" (p. 134), and she recommends that EAP programmes provide "support in English language [and] unfamiliar methodology such as problem-based learning" (*ibid*.). Thus, in terms of investigating the usefulness of EAP assessments, academic acculturation is a key area for consideration.

2.4 Is there a difference between home and international students?

Thus far the discussion has been centred on the 'issues' faced by international students and the areas in which they might be considered lacking. It has also implied that Western students are skilled critical thinkers owing to their educational background (Ryan and Louie, 2007). However, since the mid-1990s, there has been a growing body of literature which discusses the academic literacy and acculturation needs of home students. As with academic acculturation as is evident from the above discussions, though, academic literacy has many definitions.

As Lea and Street (1998) explain, academic literacy has its origins in new literacy studies, and Lea (1998) begins her definition of academic literacy by first defining literacy as being "concerned with the social practices which surround the use of the written word" (p. 158). She highlights that learning in a HE environment "involves adapting to new ways of knowing: new ways of understanding, interpreting and organising knowledge" (Lea, 1998, p. 158). A particularly important point she makes in terms of adult learners in higher education is that they "are learning to construct their knowledge in ways which make it appropriate for assessment" (Lea, 1998, p.165). Building from this, Lea and Street (1998) define academic literacy practices as "reading and writing within disciplines" (p. 158), adding that these practices "constitute central processes through which students learn new subjects and develop their knowledge about new areas of study" (Lea and Street, 1998, p. 158).

A key source which considers the academic literacy of home students and presents a useful series of definitions is the work of Bonanno and Jones (2007). In building on the original work of Bonanno and Jones (1997), Holder *et al.* (1999, p. 22) present a list of

areas in which students need to be proficient in order to cope with their HE level

studies:

A. Use of source material;
B. Structure and development of text;
C. Control of academic writing style;
D. Grammatical correctness.

More recently, Luna (2002, p. 596), citing several authors in the field, describes

academic literacy as follows:

> Going to college has been characterized as a process of initiation through which a
> student, at first an outsider, gains particular academic literacies, thus becoming a
> member of certain academic discourse communities… Inherent in this
> conceptualization is a view of literacy as ideologically situated social practices …
> and a vision of the academy as a culture that values particular ways of thinking,
> perceiving and performing.

Further, Hallet (2013) brings in the idea of wider academic literacies, which allow

students "to unpack issues of meaning and identity in relation to their chosen filed *[sic]*

of study" (p. 527), and Goldingay *et al.* (2016) explain that academic literacies require

students to build "an awareness of the discipline specific skills and complex literacy

practices … [they] need to negotiate within the university context" (p. 335).

To summarise the above definitions, academic literacy, or literacies, revolve around

students being able to negotiate the written word for study purposes. This involves

being able to read the texts required to study, but also being able to produce writing for

the purposes of assessment which is deemed 'acceptable' by the people who work in

their areas of study. Interestingly, however, Lea (2004) takes the idea of academic

literacies further than the reading and writing which students need to do to be able to

achieve at university, and she in fact calls for a "broadening [of] the contexts" (p. 743)

of what constitutes academic literacies. This broadening, I believe, links to my ideas of

assessment acculturation as explained in the previous section. Lea (2004) highlights the

importance of:

More recent literature in the areas of academic literacies has commented that, owing to changes in the student population, in particular due to the increased numbers of students from 'non-traditional' backgrounds (Bergman, 2016; Goldinay *et al.*, 2016; Lea, 2004) "there may be parallels or similarities between the learning journeys of East Asian students and those of some [home] students" (Durkin, 2008, p. 18). Goldingay *et al.*, (2016) further highlight that "recent literature has identified that some students from non-traditional backgrounds may struggle in the transition to university due to unfamiliarity with language and unwritten expectations at university" (p. 334). Further, Cumming *et al.* (2016) also comment that in the studies they examined, "the universal nature and challenges of learning to write effectively from sources makes it difficult to draw absolute distinctions between the writing of first and second language students" (p. 52). In terms of the deficit model as outlined above, then, the studies discussed so far in this section demonstrate that it is not *just* international students who need support with their transition to university. In fact, as Grabe and Zhang (2013) explain, "learning to write from textual sources (e.g. integrating complementary sources of information, interpreting conceptually difficult information) is a challenging skill that even native speaker students have to work hard to master" (p. 9), with Lea (2004) commenting that often *all* students experience difficulties "when trying to unpack the ground rules of writing in any particular context" (p. 749).

It is clear from the above discussions, and those on the criticisms of the deficit model presented previously, that both home students and international students have similar issues with their adaptation to the new institution. But I believe that since, as explained

previously, international students are coming from, at times, very different educational backgrounds, the adaptation required by them is much larger. Take, for example, the details provided by Pierson (1996, p. 55), which, although an older source now, describes the Chinese education system thus:

> The present colonial education system, with its excessive workloads, centralized curricula, didactic and expository teaching styles, concentration on knowledge acquisition, examinations emphasizing reproductive knowledge over genuine thinking, overcrowded classrooms, and inadequately trained teachers.

Further, Littlewood (2000) comments that the passivity of international students as highlighted in the previous section is most likely a result of their previous educational experiences. Thus, in order to further aid international students with their transition to the new academic culture, and to support them with their academic literacy development, it is important that they are exposed to institutional norms and requirements as part of the EAP programmes. Further, although the deficit model is not the most appropriate way of viewing international students, nor is it a model which I wholly agree with, Cumming *et al.* (2016) do report the findings of several studies which have shown that "second-language learners had greater difficulties formulating appropriate task representations, demarcating their scope and intentions, and deciding on strategies for writing than did their native English-speaking counterparts" (p. 51). This implies, then, that to some extent international students do have gaps in their knowledge and understanding in terms of their academic literacies when compared to home students.

Considering the above points, then, the academic acculturation of students, therefore, should be an integral part of any EAP programme, and this should be reflected in its corresponding assessments, particularly since, as will be discussed below, these can be used to help develop not just language but additional academic and personal skills. Essentially, as Hyland and Hamp-Lyons (2002) explain, EAP should involve grounding

teaching and assessment in "an understanding of the cognitive, social and linguistic demands" (p. 2) of the new academic environment with a view to "equipping students with the communicative skills to participate in particular academic contexts" (*ibid.*). This will aid students with their adjustment, since, as Kember (2000) explains, international students can, and do, adapt, even if only temporarily, if staff "take steps to ease the transition process" (p. 111). He adds that "initial guidance and explanation is important, and teachers should be prepared to provide support" (Kember, 2000, p. 118). Since many international students see the EAP programmes as a "training ground" as explained above, it seems that EAP classes and assessments are key spaces in which to help students with this complex academic adjustment.

2.5 What is meant by 'assessment'?

Before coming to language assessments, I will first take a more general view of assessment, considering in particular assessment in HE, which Hernández (2012) views as "central to the students' experience" (p. 489). The importance of assessment in HE is picked up by several other authors too, but before getting to that, I will offer a few definitions of assessment as discussed in the literature. Starting with Price *et al.* (2011), their definition of the purpose of assessment explains that it has several functions: "assessment needs to motivate and challenge the learner, to stimulate learning and to provide feedback" (p. 481). In a more recent article, Lau (2016) explains that assessment should contribute to student learning, commenting that in order for it to do so, firstly, formative and summative assessment need to be connected with each other, and secondly assessment as a whole needs to be connected with the overall learning and teaching environment. In summarising various key European documents relating to pan-European education, Pereira *et al.* (2016, p. 1009) highlight that

> assessment requires students to be evaluated through explicit criteria for marking,
> with procedures adequate for formative, summative or diagnostic assessment
> purposes, and the assessment strategy clearly communicated to students. In

addition, assessment is an indicator of teaching and learning effectiveness, and its outcomes impact on the future careers of the students.

Citing further sources, they emphasise the European requirement that there should be "a student-centred learning approach that helps students develop competences needed in real life" (Pereira *et al.,* 2016, p. 1009).

In discussing the purposes of assessment, Walklin (1990) explains that assessment can be used to assess individual student attainment, assess the progress of a particular group or to discover student learning problems. This is also commented on by Dunn *et al.* (2004), although they present a more comprehensive list of the roles and purposes for assessment:

- Diagnosing student difficulties
- Measuring improvement over time
- Motivating students to study
- Judging mastery of essential skills and knowledge
- Ranking the student's capabilities in relation to the whole class
- Evaluating the teaching methods
- Evaluating the effectiveness of the course
- Encouraging the tacit learning of disciplinary skills and conventions

Dunn *et al.* (2004, p. 16)

Fletcher *et al.* (2012, p. 119) also believe that HE assessment is multifarious in nature, and highlight the contradictions surrounding it:

Assessment is meant to inform student learning even as it sorts students into those who pass and those who fail; assessment measures learning outcomes but also compares students with one another; assessment should be objective and individually accountable but must evaluate the attainment of dispositions such as creativity, leadership and imagination.

To return now to why assessment is so important, Wakeford (2003, p. 42) explains that

assessment is an integral component of the teaching and learning system. Assessment may be used explicitly to guide students in their study. But also, student perceptions of what is rewarded and what is ignored by more formal examination procedures will have a substantial impact upon their learning behaviour and thus upon the outcomes of a course.

This importance of assessment is commented on by several other authors, and Boud *et al.* (1999) "assessment is the single most powerful influence on learning in formal courses and, if not designed well, can easily undermine the positive features of an important strategy in the repertoire of teaching and learning approaches" (p. 413). They also comment that "assessment exercises power and control over students" (Boud *et al.*, 1999, p. 418), adding that assessment also "exerts a backwash effect on learning" (*ibid.*). Taras (2005 and 2008) emphasises the central importance of assessment in education, stating in particular that it "has been shown to be the single most important component that influences student learning and education in general" (Taras, 2008, p. 389). Further, Bloxham and Boyd (2007) explain that assessment in higher education in particular "shapes the experience of students and influences their behaviour more than the teaching they receive" (p. 3), and Price *et al.* (2011) comment that "the majority of students regard assessment as the most important aspect of their course, and consequently use it to guide and frame their learning" (p. 480).

Despite the obvious centrality of assessment to the learning experience of students, however, Taras (2005) comments that "there is a lack of commonality in the definitions of the terminology relating to it" (p. 466), and so I now wish to investigate some of the key definitions linked to assessment. There are various types of assessment which are used in HE contexts in the UK. These include formal examinations, essays, reports, dissertations, seminars, presentations, debates and even practical examinations (Brown and Knight, 1994; Clegg and Bryan, 2006; Boud and Falchikov, 2007; Norton 2007). These assessments, however, are complex in nature in that they serve more than one purpose. In fact as Pereira *et al.* (2016) assert assessment in higher education has changed considerably in recent times, citing that assessment approaches are now "more focussed on students and learning" (p. 1008).

The two key terms which are returned to throughout discussions on assessment are formative and summative, and no discussion relating to assessment in any educational context can avoid these terms, so I will now discuss what these terms mean, with a particular focus on HE. Starting with Dunn *et al*. (2004), they explain that summative assessment involves tasks which "measure the sum of [student] performance" (p. 18), whereas formative assessment allows "students to improve their performance *on the same task*" (*ibid*, original emphasis). In providing definitions of the two, Crisp (2012) explains that "the two most common types [of assessment are] formative (designed primarily to improve learning) and summative (designed primarily to judge learning)" (p. 33). These are quite simplistic views, however, and present what Price *et al*. (2011) call a 'binary view' of assessment, highlighting the often dichotomous nature (Lau, 2016) of discussions about different approaches to assessing students. In fact to further problematise the issue, Black and Wiliam (1998) explain that "the term formative assessment does not have a tightly defined and widely accepted meaning" (p. 7). They, however, choose to define it as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (pp. 7 – 8).

In examining the issues of the binary or dichotomous approach, Price *et al*. (2011, p. 484) comment that

> for many, the distinction between summative and formative assessment is clear: summative assessment generates marks and regulates whether students can pass through a specific boundary when moving towards accreditation. Formative assessment, on the other hand, gives students information on how their learning is progressing.

However, they go on to comment that an assessment can serve both purposes, it can be both summative in that it is used to decide student achievement and it can be formative provided it offers students feedback on how they can improve (Price *et al*., 2011). This

supports Black and Wiliam (2003) who explain that "it has long be recognised that assessment can support learning as well as measure it" (p. 623). Lau's (2016) review article about formative and summative assessment presents additional discussion of the dichotomous nature of the two terms. In reviewing the origin of summative and formative as ways of describing assessments, she emphasises that although originally seen as complementary forms of assessment, with formative introduced as "a way to improve students' summative performance" (Lau, 2016, p. 512). She further explains that older research in the areas of assessment emphasised that formative and summative assessment should be "linked together" (*ibid*.). Thus, rather than favouring one approach over the other, what Lau (2016) recommends is "a balance between the two models" (p. 516). These points further complicate the issues of assessment, and offer further evidence to support the need to offer formative feedback in summative assessments.

As well as the issues of terminology, it is also highlighted in a number of sources that assessment in HE has been in a state of flux since the 1990s (Knight, 2002; Black and Wiliam,1998; Hager and Butler, 1996). In terms of the changes, firstly there have been changes to the ways in which students are 'measured'. Hager and Butler (1996) explain that from the 1970s, educational assessment "has been dominated by a model of scientific measurement" (p. 367), adding that from the mid 1990s, this traditional approach has been challenged, with a move towards what they call "the judgemental model" (p. 367). What this means is that more recent developments in educational assessment mean that students are being measured less and less through standardised tests, and applied knowledge is being assessed using approaches such as essays, reports and portfolios. Further, Black and Wiliam (1998) comment that during the 1990s, there was "a shift in focus of attention, towards greater interest in the interactions between

assessment and classroom learning and away from concentration on the properties of restricted forms of test which are only weakly linked to the learning experiences of students" (p. 7).

What constitutes these 'new' or 'alternative' approaches to assessment is summarised by Pereira *et al.*, (2016) as being focussed around problem-based and case-based assessments, as do self- and peer- assessment. Case-based and problem-based assessments are reported as "promot[ing] the development of professional skills and learning in real-life contexts" (Pereira *et al.*, 2016, p. 1010). Self-assessment is defined as "a student-centred approach that engaged learners in the learning process, promoting feedback and developing the students' ability to learn and to self-assess" (Pereira *et al.,* 2016, p 1023). The effectiveness of peer-assessment is highlighted as being due to it "allow[ing] interaction between students … and produc[ing] formative feedback" (*ibid.*).

In terms of assessment in general, then, a key area for discussion in a great deal of the literature which relates to assessment in HE is the call for a move away from the more formal, summative assessments which more often than not come at the end of a module, unit or programme of study, to a more formative approach. Although Carless (2009) was of the opinion that "examinations maintain their position as a major form of assessment in higher education" (p. 82), several other authors (Clegg and Bryan, 2006; Price and Donovan, 2006; Boud and Falchikov, 2007; Norton 2007) have commented on the move to the more formative approach based on the belief that summative assessment has only a limited impact on learning and can lead to the strategic allocation of time and effort (Gibbs, 2006; Crisp, 2012). This is reflected by Lincoln and Kearney (2015) who comment that "sound assessment instruments need to be based on strategies

that go beyond aptitude tests or predictors, graduate surveys, and university rankings"

(p. 391). Thus, the literature suggests that assessment in HE, then, seem to be moving

towards a more ongoing, formative approach, offering chances for development, rather

than merely testing knowledge at the end of a semester or programme (Fletcher, *et al*.,

2012).

Essentially, it is believed that HE assessment should be developmental. Although at

first glance perhaps a simple distinction, Hernández (2012) highlights that the difference

between formative and summative assessment, is not as clear cut as some authors might

believe. In fact she explains:

> Some assessments in higher education are designed to be both formative and
> summative ... because they provide feedback so that the students can learn from it.
> Furthermore, the same assessment task fulfils a summative function because a
> grade is awarded and it contributes to the overall results of a course.
>
> Hernández (2012, p. 490)

Barnett (2007) agrees with this, and states that summative assessment does, in fact, have

"a definite role to play in assisting the student's development" (p. 29). Additionally,

Fletcher *et al.* (2012) highlight that assessments "provide information to faculty about

teaching effectiveness … and to students about how well they are doing and how they

can improve their learning" (p. 119).

In terms of assessment in HE, however those who design the assessments find

themselves under increasing amounts of pressure from a number of directions. Firstly,

there are the students, who continually rate assessment lower on the National Student

Survey each year (Hefce, 2016; Price *et al.,* 2011). Additionally, there is increasing

pressure from institutional management who are attempting to respond to the low

student satisfaction levels, often by reducing the amount of assessment used on

programmes of study (Price *et al.,* 2011). Further pressures from industry and

employers who "continue to seek graduates who can contribute to [the] competitive business sector" (Clements and Cord, 2013, p. 114). Indeed, Hughes' (2013) study about graduate learning outcomes also highlights the fact that learning outcomes tend to "focus on employability at the expense of transformational outcomes" (p. 493), with Knight (2002) also commenting on the facts that HE institutions in the UK "are now accountable for student employability" (p. 278). Finally, there are the academic pressures of being able to make judgements about student performance in order to be able to grade them (Lau, 2016; Price *et al.,* 2011; Bryan and Clegg, 2006; Black and William, 2003). Lau, in fact highlights that "external pressures have led to lecturers' efforts being increasingly directed towards summative assessment for certification and quality assurance purposes" (Lau, 2016, p. 512). Add to this the increased pressures of limited time and increased student numbers (Price *et al.,* 2011) and those who design assessments find themselves having to balance innovation against practicality, skills development against knowledge testing, and the needs of students against those of management and employers. Thus, despite the continued developments in HE assessment practices as reported by authors such as Lau (2016), Pereira *et al.* (2016), the 'standard' forms of assessment such as exams, essays and reports continue to dominate in higher education.

It is clear from the above discussion, then, that assessment in higher education is a complex business, and the assessments serve various purposes. Further, assessment appears to be changing, moving from taking a purely summative role, to being formative and summative at the same time. The decisions about how to assess are often made taking into account the needs of various interested parties, including quality assurance agencies, employers and the students themselves. Purely assessing what has been taught to see if it has been learnt is no longer sufficient, and with this in mind, the

definition of assessment in HE which I will use is those activities which are used at the end of a unit, module or programme of study, which can play both a formative and summative role, and which impact on both teaching and learning activities. These assessments must also offer information for employers as to the wider abilities of students, and must satisfy those who work in quality assurance as being fair and just measurements of ability.

2.6 What is a language assessment?

Having outlined briefly the history of EAP and provided a working definition of the term for the purposes of this thesis, as well as considering the links between EAP and academic acculturation and discussing assessment in HE, the next point for discussion is to establish what it is that constitutes a language assessment. With HE assessment in general being such a complex term, it follows that defining 'language assessment' will be equally difficult, and despite the relatively short history as an academic discipline alluded to at the start of this chapter, Donohue and Erling (2012) explain that "language assessment is a highly developed field of activity that has given rise to a wealth of research" (p. 211).

Heaton (1975), and later Douglas (2010) have both offered relatively straightforward definitions of a language test. Combining their definitions, this study takes a language assessment in its most basic form to be an instrument with which to measure language ability and evaluate language performance (Heaton, 1975 and Douglas, 2010). This is, however, a very simplistic view, and as McNamara (2000) explains, despite falling under this broad definition, tests do, in fact, normally differ in two key ways: test method and test purpose. He further explains that there are two main methods of testing a language: pencil and paper tests, which "take the form of the familiar examination

question paper" (McNamara, 2000, p. 5) and performance tests, "in which language skills are assessed in an act of communication" (*op. cit.* p.6). To return to the discussion of EAP offered above, then, the definition of an EAP assessment is complicated since the assessments being investigated here span both of these definitions depending on which assessment the students are doing.

Taking this into consideration, the point of departure for any evaluation of language assessments is to establish which test method is being used. There are, however, various additional definitions from key literature on the subject of language assessments (for example Douglas, 2010; McNamara, 2000; Jordan, 1997; Davies, 1990; Madsen, 1983; and Davies, 1977) which need to be further explored in order to provide a clearer answer to the question posed above, and the starting point is actually another question, one propounded by Heaton (1975): 'why test?'. In order to answer this seemingly simple question it is necessary to problematize testing further.

The starting point for problematizing testing is the terminology that we use. This is touched on by Jordan (1997) who explains that "the words 'test' and 'examination' are often used synonymously" (p. 85). He adds that:

> sometimes a distinction is made according to the criterion of formality or external/public, where 'test' is smaller, local, less formal and often part of an ongoing process of evaluation or feedback.
>
> Jordan (1997, p. 85)

As can already be seen, then, there is a great deal of synonymous terminology used in the area of assessment, so for the purposes of this study, I will be using the terms 'exam' and 'assignment', the distinction being that 'exam' is a one-off, end-of-programme assessment, whereas 'assignment' is a longer piece of coursework such as an essay or a report, which is submitted at the end of a longer period of time.

Additionally, 'assignment' will be used to refer to assessments such as seminars and presentations, for which students have upwards of a week to prepare. I consider all of these should come under the umbrella term 'assessment' which I take to mean any activity which is used for evaluating student performance and assigning a grade.

In terms of types of test, Davies (1977, p. 45) is a key source for the definitions of language assessment. He explains that there are two main types of assessment: achievement tests, which are a way to assess what has been learned of a known syllabus, and proficiency tests which are "concerned with assessing what has been learned of a known or unknown syllabus" (Davies, 1977, pp. 45 – 46). Madsen (1983) further expands these definitions, adding that achievement tests "simply measure progress" (p. 9), whereas proficiency tests "measure overall mastery of [English] or how well prepared [the student] is to use [English] in a particular setting" (p. 9). These, however, are quite old, and some may say outdated, definitions of language assessment, and for the purposes of this study, they need to be expanded since the assessments in question are EAP assessments and fall more into the communicative language assessment definition. Canale and Swain (1980, p, 34) in their influential article on the subject of communicative approaches to language teaching highlight that:

> communicative testing must be devoted not only to what the learner knows about the second language and about how to use it (competence) but also to what extent the learner is able to demonstrate this knowledge in a meaningful communicative situation.

In the case of the assessments being investigated here, the knowledge about English and how to use it needs to be demonstrated in an academic situation. Hence, a communicative approach is the most appropriate.

With these ideas in mind, Davies (1990) explains that "human, cultural knowledge does not reduce to mere number" (p. 1), and he provides a clear definition of a language testing, explaining that "it represents a measured concentration on language use and knowledge" (Davies, 1990, p. 9). He also explains the varied uses of language tests, pointing out that they can be used "for selection, for feedback and, less frequently, for evaluation and research" (p. 9).

To summarise this section, then, I return to Douglas (2010) who explains that "in a language test, we direct language input at test takers and infer qualities of language ability based on the outcome, or language performance" (p. 18). In the case of this study, the language input being directed at the test takers is more usually academic in nature, and the qualities of language ability being inferred based on this outcome needs to be an estimation of the students' ability to operate and use language successfully in an academic context.

2.7 Why do we use language assessments?

As is clear from the above discussions, assessments have various functions, although in the field of language, especially language for specific purposes, they may in fact be multifaceted serving several purposes simultaneously. An English writing assignment might, for example, be used to diagnose a student's difficulties in various areas of language use (grammar, syntax, organisation) whilst at the same time judging their mastery of these and other skills (reading-for-writing, paraphrasing, effective use of references) as well as encouraging the tacit learning of disciplinary skills and conventions which could be applied to other assignments once students move on to their main course of study. So, since we are considering English language assessments here, this section will focus on the reasons which lie behind language testing.

Davies (1990) outlines what he calls the "*triple message* that language assessment always provides" (p. 10, original emphasis). This triple message includes: a message about skill, a message about development and a message about knowledge. In outlining the deeper meaning of this triple message, he explains that the message about skill refers to "the extent to which learners have reached adequate proficiency" (Davies, 1990, p. 11); the message about development "provides an indication – through assessment – as to the psycholinguistic and the sociolinguistic provenance of the learner" (Davies, 1990, p. 11). Finally, the message about knowledge which is shown through

> the range of acceptability judgements [students] are prepared to make and the extent of their conscious metalinguistic reflecting upon language, which in turn demonstrates itself in knowledge about language and in areas of ludic creativity.
>
> Davies (1990, p. 11).

This final message is linked closely to the message about skill, since, as Davies (1990) explains, "there may well be an element of knowledge within skill which determines differential proficiency" (p. 11). The above definitions suitably define the EAP language assessments being used at the case study HEI and already it is clear that language assessment is a complicated process when we attempt to take it apart to investigate, yet there is a further complication, which is considered by Douglas (2010).

Having outlined the basic reasoning for why language assessments are used, I will now return to discussing the two main types of assessment: achievement and proficiency. Bachman (1990), McNamara (2000) and Douglas (2010) all offer useful definitions of the differences between achievement and proficiency tests which demonstrate why language assessments are actually used. Bachman (1990) explains that proficiency tests are theory-based, whereas achievement tests are syllabus-based. McNamara (2000)

adds to this, explaining that achievement tests "are associated with the process of instruction" (p. 6), but that proficiency tests "look to the future situation of language use without necessarily any reference to the previous process of teaching" (p. 7). Douglas takes these ideas further, highlighting that

> achievement tests are based on the actual material learners have been studying or upon commonly agreed understandings of what students should have learned after a course of study. [...] Proficiency tests are usually intended to help us make predictions about what learners will be able to do with the language in communicative situations outside the classroom.
>
> Douglas (2010, pp. 1-2)

In an earlier work, Davies (1968) provided useful definitions of the various types of language tests and their uses. He also explained that "there is some confusion in the use of terminology in Language Testing" (Davies, 1968, p. 6), adding that the main confusion seems to be between the terms achievement, attainment, proficiency, diagnosis and aptitude. To help clarify things further, he provided some useful definitions of these terms as reproduced in Table 2 below:

Table 2 – Types of Language Testing: Adapted from Davies (1968, pp. 6 – 7)

Achievement	Proficiency	Diagnosis
Also known as attainment. This type of test assesses the student's knowledge 'as indicative of past learning' (Davies, 1968, pp. 6 -7). The achievement test is used to find out how much has been learnt of a known syllabus. It should not, however, be used to make predictions about students' performance in the future.	Also known as aptitude. This type of test is used to assess 'the student's present ability for future learning' (Davies, 1968, p. 6). The proficiency test is 'concerned with both the past and with the future' (ibid. p. 7), and it is used 'to predict skill in language for some extra-linguistic purpose' (p. 7), for example, a student's ability to cope in UK HE. An aptitude test, on the other hand, does differ slightly in that instead of looking at the extra-linguistic purpose, it focuses on a particular language skill, and unlike proficiency, aptitude has no interest in past learning.	A diagnostic test is used by the teacher to assess 'what has gone wrong' (Davies, 1968, p. 7). Diagnosis differs from other tests in that it 'relates entirely to the use made of the information and not at all to the presence of a skill in the learner' (Davies, 1968, p. 7). Unlike achievement or proficiency, diagnostic testing is entirely for teacher use, it is not often used to inform the learner of their skills.

To summarise the various definitions then, achievement assessments evaluate how well the learners have learnt what they have been taught, whereas proficiency assessments focus on the learners' potential ability to use the language in future situations. As was explained in the previous sub-section, the use of the language assessments being investigated in this study is not that clear cut, as Table 2 implies. Further to this, the assessments under investigation are based on the actual material which learners have been studying, making them achievement tests, but they are also used to help make predictions and inferences about what learners will be able to do with the language when they move on to their main programmes of study (Douglas, 2010 and McNamara, 2000), meaning they are also proficiency tests. This is a point which has not been discussed in much detail in the EAP literature considered so far since the majority of it is focussed on course content and teaching methods rather than on the how and what of assessment. This how and what of EAP assessment will be investigated in Chapter 3.

2.8 Summary

In the process of reviewing the literature so far, some important generalisations have emerged:

1) EAP is complex to define in neat terms since skills are often integrated and tests serve many purposes, including language development, skills development and academic acculturation; the extent to which all of this should, or even can, be included in a battery of assessments merits further study.

2) Literature relating to academic acculturation and international students often presents students from a deficit point of view. Despite this, literature related to academic literacies indicates that all students might be viewed in deficit. Despite this, research also implies that the issues faced by international students are further exacerbated by the language issues they encounter, therefore

investigating the merits of including academic skills development and acculturation in EAP assessments is an interesting area for investigation.

3) HE assessment literature indicates that assessment in HE in general has been evolving for some time, and in more recent years there has been a move away from the more traditional forms of assessment such as examinations or essays. This has potentially large impacts on the way EAP is assessed and thus current HE assessment approaches need further examination.

4) It is often recommended that EAP course design be based on a detailed needs analysis, yet the same does not seem to have been recommended for EAP test design; this is an area which also warrants further investigation.

In completing this chapter, I have briefly outlined the history of EAP, and have reviewed some of the key literature related to EAP, its link to academic acculturation, assessment in general and language assessment. In doing this, I have achieved the first two objectives of this study. The three generalisations above indicate that further research into these areas is required in any discussion of EAP assessment usefulness. Essentially, the complexities with EAP mean that the design of any assessments which are used to assess whether students have acquired the skills necessary as well as developed the required language competence to be able to handle the various academic tasks which they will encounter when they progress onto their main programmes of study is equally complex in nature. In order to evaluate any battery of EAP assessments to establish their usefulness, an evaluative framework needs to be designed, which is the purpose of the following chapter; this will help achieve the third objective of this study.

CHAPTER 3

DEVELOPING AN EVALUATIVE FRAMEWORK

3.1 Introduction

Thus far I have provided some background to what EAP is, as well as some general discussions of assessment and language assessment. This has outlined the complexities of the area being discussed. The generalisations arising from the literature reviewed in Chapter 2 demonstrate areas which warrant further study, and in order to study these areas, a clear evaluative framework is required. So as to establish this, I will answer three further questions: Why does good test design matter? How can we assess EAP? and How can we evaluate EAP assessment usefulness? This chapter will provide further review of literature, and then present an evaluative framework for application in the second part of this thesis. To do this, the first section will consider methods for EAP assessment, considering what constitutes academic language ability and how this can be demonstrated. The second section will present a discussion of the models and theories proposed by the various authors for evaluating the tests being considered here. In doing this, I will achieve the third research objective for this study: To develop an evaluative framework to establish the usefulness of EAP assessments.

3.2 Why does good test design matter?

The consequences of test use is an area which many authors have considered, and is an area which has been the cause of much disagreement in the literature. Chalhoub-Deville (2016) explains that these consequences are discussed in the literature under several different headings depending on the discipline or even the geographical orientation of the writer. The terms Chalhoub-Deville (2016) is referring to are 'impact', 'consequences', 'backwash' and 'washback'. How these consequences should be studied and discussed is, however, the main area of disagreement in the literature.

Some authors (see, for example Messick, 1989; Bachman and Palmer, 1996; Lynch, 2001; Brennan, 2013; Kane, 2013a; Chalhoub-Deville, 2016) believe these consequences should be considered as part of the validation argument (see section 3.4.2 below for further discussion of this), whereas others imply that including test consequences in a validation study is not feasible or even necessary (Popham, 1997; Shepherd, 1997).

There are, in fact, many reasons why it is important to have a well-designed language test, more of which will be discussed as this chapter progresses, but this first section will consider just one of these areas: washback. Test washback is not a phenomenon which is exclusively seen in language assessment, but it is certainly a key consideration in evaluating language test usefulness. Alderson and Wall (1993) explain that "the notion that testing influences teaching is commonplace in the educational and applied linguistics literature" (p. 115). In defining washback, Green (2013) explains that it is "the impact that a test has on the teaching and learning in preparation for it" (p. 40), with Cheng and Curtis (2010) defining it as "the influence of testing on teaching and learning" (p. 3). Essentially, assessments can impact on syllabus design and teaching methodology: a useful test tests what is taught, but equally a good (or indeed bad) test impacts on what is taught and how.

Looking at washback in terms of language testing, Bailey (1999) explains that "tests exert a powerful influence on language learners who are preparing to take these exams, and on the teachers who try to help them to prepare" (p. 1). Although Bailey is talking about a large-scale, high-stakes assessment, in this case the TOEFL exam, the same can be said for the assessments being considered for this study; they may be small-scale, but they still have a considerable influence on the teachers and the students, be it good or

bad. Alderson and Wall (1993) present an interesting study titled 'Does washback exist?' explaining that washback is "common in language teaching and testing literature and tests are held to be powerful determiners of what happens in the classroom" (p. 117). Malone and Montee (2014) agree, stating that "washback is commonly defined as the changes that occur in teaching and learning as a result of test use" (p. 4). Bachman (1990) holds a similar opinion and highlights that this effect on what happens in the classroom can be either positive or negative, with positive washback being the result of "the testing procedure reflect[ing] the skills and ability that are taught in the course" (p. 283). He also explains that it could be negative, "where there is little or no apparent relationship between the types of tests that are used and the instructional practice" (Bachman, 1990, p. 283).

Alderson and Wall (1993) also discuss negative and positive washback, although their take on these terms is somewhat different to that of Bachman (1990) or Bachman and Palmer (1996). They believe that if a test has negative washback, then it could be that teachers may "ignore subjects and activities which [do] not contribute directly to passing the exam" (Alderson and Wall, 1993, p. 115). In a brief summary of sources about negative washback, Qi (2011, p. 221) explains that:

> negative washback effects include the facts and observations that the relevant test, rather than the curriculum, dictated what skills and knowledge were taught … that textbooks became clones of the test … that tests discouraged students from putting genuine effort into their learning … and that a test reduced teachers' abilities to teach in ways that are incompatible with standardized testing formats

These may be issues with standardised tests, but I would argue that if tests are designed with a specific purpose in mind, take into account an institutional context and are designed with a view to helping students develop language and skills they need to cope within a certain setting, then washback is not always negative. Qi (2011) in fact discusses this area, highlighting that positive washback is "demonstrated when the tests

motivated students to work harder, or made them aware of new areas of language" (p. 221).

Interestingly, however, Alderson and Wall (1993) also explain that *any* test might have positive or negative washback, emphasising that a good test can have bad impacts if it limits teaching and learning, whereas a bad test could have good impacts if it influences "teachers and learners to do 'good' things they would not otherwise do" (p. 117). This idea is supported by Messick (1996), who comments that "a poor test may be associated with positive effects and a good test with negative effects because of other things that are done or not done in the educational systems" (p. 5).

If Estaji and Tajeddin (2012) are correct when they say that washback is "grounded in the relationship between preparation for success on a test and preparation for success beyond the test, in the domain to what the test is intended to generalize and to which it may control access" (p. 6), then the washback effect of tests must be considered when establishing assessment usefulness. Thus, since testing can have such an impact on syllabus design, lesson planning, teaching approaches and student learning, it is vital that what is assessed in EAP at the very least supports students with their language skills improvement as well as their academic acculturation and academic skills acquisition. Fulcher (2010) provides a bullet-pointed list stating what a test will influence:

- What teachers teach
- How teachers teach
- What learners learn
- How learners learn
- The rate and sequence of teaching
- The rate and sequence of learning
- Attitudes to the content, method, etc. of teaching and learning

Fulcher (2010, p. 278)

This idea of washback and how students are taught and learn are areas I will return to when I discuss target language use situations and test task authenticity in sections 3.4.1 and 3.4.2 below. How to decide what to test, and more importantly how to test this, are however the first areas for discussion when evaluating EAP assessment usefulness, and the next section will consider how to go about assessing EAP.

3.3 How can we assess EAP?

When designing effective EAP assessments, the specific needs and goals of the institution and departments the students are moving into need to be defined (Ypsilandis and Kantaridou, 2007). It is only by defining the educational norms and cultures of these that we will know what skills are needed and how best to assess them. Taylor and Geranpayeh (2011) explain that the challenge which faces anyone trying to develop an academic language assessment is that they need to "ensure that the cognitive processing activated in the test taker by a test task corresponds as closely as possible to what they would expect to do in the academic … context" (p. 96). Linked to this, Ferris and Tagg (1996a, 1996b) find that the relative importance of the types of language task (for example taking notes in lectures, giving presentations, writing essays, reading extended and possibly technical texts) varies considerably according to both programme and level of study. This relates back to the acculturation discussion above in section 2.3. Further, it is for these reasons that I am using a case study institution as these should be easier to define[4].

As Bachman and Palmer (1996) explain, "the way we define language ability for a particular testing situation, then, becomes the basis for the kinds of inferences we can make from the test performance" (p. 66). Therefore, deciding just how to assess EAP,

---

[4] I will return to this point in the Methodology in Chapter 4.

essentially, defining what it is that we are looking for, is vital since it will form the basis of the start of the evaluative framework. A starting point for this is the work of Tajino *et al.* (2005) in which they ask a series of questions about tests. I have adapted two of these into questions to be used as a starting point for thinking about how to assess students in EAP:

1) How can we effectively integrate academic skills testing and test development and language testing and test development?

2) How can we ensure that students have developed the level of proficiency they need to be able to cope with HE level study at MUC?

These questions reflect Bachman and Palmer (1996, p. 67), who state that "there is a need to be aware of the full range of components of language ability as we design and develop language tests and interpret language test scores". It is this full range of components of language ability that is covered by the first of the questions; the second question will be expanded further as this section progresses.

In assessing language ability, in this case academic language ability, Bachman (1990) proposes that first it is necessary to establish what language ability actually is. His review of the various definitions is rather extensive, however based on this, he (*ibid.*) puts forward a model of communicative language ability as applicable to the EAP context in which he outlines three key components:

1) Language competence

2) Strategic competence

3) Psychophysiological mechanisms/skills

Bachman (1990), in fact, highlights that a common thread in language learning and

testing research of the late 1980s was:

> the belief that a precise, empirically based definition of language ability can
> provide the basis for developing a 'common metric' scale for measuring language
> abilities in a wide variety of contexts at all levels and in many different languages.

> Bachman (1990, p. 5)

However, as Bachman and Palmer (1996, p. 66) explain:

> if we are to make inferences about language ability on the basis of performance on
> language tests, we need to define this ability in sufficiently precise terms to
> distinguish it from other individual characteristics of test performance.

Demonstrating how "performance on [a] language test is related to language use in

specific situations other than the language test itself" (Bachman and Palmer, 1996, p.

10) is the crux of this thesis since my aim is to establish how useful language tests are.

Bachman and Palmer (*op. cit.*) add that in order to be able to do this, what is needed is a

conceptual framework that will enable lecturers or test users to "describe what [they]

believe are the critical features of both language test performance and non-language test

use". It is this conceptual, or in this case evaluative, framework that will form the basis

of the next chapter.

In deciding how to assess EAP, what is needed is a more general definition of *academic*

language ability as presented earlier in this chapter. This definition, however, is more

complex than it at first seems since language use for academic purposes is highly

integrated and thus needs further discussion.

Pot and Weidemen (2015) for example, present a list of ten components which make up

academic language ability:

1) Understand a range of academic vocabulary in context
2) Interpret and use metaphor and idiom, and perceive connotation, word play and ambiguity
3) Understand relations between different parts of a text be aware of the logical development of (an academic) text, via introductions to conclusions, and know how to use language that serves to make the different part of a text hang together
4) Interpret different kinds of text type (genre), and have sensitivity for the meaning that they convey and the audience that they are aimed at
5) Interpret, use and produce information presented in graphic or visual format
6) Make distinctions between essential and non-essential information, fact and opinion propositions and arguments; distinguish between cause and effect; classify, categorise and handle data that make comparisons
7) See sequence and order, do simple numerical estimations and computations that are relevant to academic information, that allow comparisons to be made and that can be applied for purposes of an argument
8) Know what counts as evidence for an argument, extrapolate from information by making inferences, and apply the information or its implications to other cases than the one at hand
9) Understand the communicative function of various ways of expression in academic language (such as defining, providing examples, arguing)
10) Make meaning (e.g. of an academic text) beyond the level of the sentence

Pot and Weideman (2015, p. 27)

Clearly, this is a very comprehensive list, and other authors provide similar lists and tables. Guse (2011) for example presents a comprehensive framework for all four language skills, as well as outlining a framework for vocabulary and grammar development; Jordan (1997) lists various study situations and activities and the study and language skills needed to achieve these. The lists are not exhaustive, and need to be complemented by the institutional context and specific institutional or departmental requirements. Saying that, it is possible to broadly define key activities which students will undertake, as well as identify a threshold language ability for achieving these in an academic context. In doing so, the test can be considered useful to all stakeholders.

3.4 How can we evaluate EAP assessment usefulness?

Having reviewed key literature relating to EAP, assessment and academic acculturation, this final section will critically review the literature related to assessment development and evaluation being used to underpin the evaluative framework that can be used to evaluate in-use EAP assessments and their corresponding grading criteria. In doing so, I will achieve the third objective of this study. Before doing this, however, some discussion of why establishing test usefulness is necessary will be presented.

Language testing, as is evident from the discussions above, is a complex issue, and it is further complicated by its paradoxical and often multifaceted nature. This is discussed by Davies (1990) who explains that, owing to the impossibility of testing a whole language, the samples of language assessed within a language test are "both totally inadequate and totally adequate at the same time" (p. 3), hence the need to ensure that they are as useful as possible. Douglas (2010, p. 18) adds to this and provides good reasoning for why establishing the usefulness of a test is needed:

> because we cannot otherwise physically encounter language ability directly, we are forced to rely on *indirect* measurement, and all language tests are in fact indirect measures of language ability.

This means that language tests provide simply the basis for inferences about language ability, rather than definitive measures of it. The measurements are indirect, and the grades awarded are inferences since it is not possible to truly predict exactly how a student will perform in a 'real life' situation (Bachman, 1990; Bachman and Palmer, 1996; McNamara, 2000). Donohue and Erling (2012) also touch on this point, explaining that language which is produced for the purposes of a language assessment is only evidence of a student's capacity to read and write in that context and at that time, rather than predicting how they will perform thereafter. Whether or not these inferences are reliable depends entirely on whether the test is a suitable measure of the abilities it

purports to be measuring; in other words, what needs to be established is whether or not the test is useful.

Bachman (1990, p. 1) outlines some useful questions in the introduction of his book on language testing. These are: "What types of test should we use?" "How long should the test be?" "How many tests do we need to develop?", adding that, to these questions, there are "no clear-cut, absolute answers". Subsequent questions which he adds to these first three are "how reliable should our tests be?" "Are our scores valid for this use?" and "How can we best interpret the results of our test?". These six questions are key when attempting to establish a framework with which to evaluate the language assessments in use at any institution, and they are a useful starting point for the development of the framework for use within this thesis. Essentially, then, while attempting to answer these questions about the language assessments being used for this study, I have attempted to demonstrate how to establish the level of usefulness of these assessments. Bachman's (1990) study is, therefore, a key text for this thesis in that it presents "a conceptual framework for answering practical questions regarding the development and use of language tests" (Bachman, 1990, p. 1) which is based on three broad areas: "(1) the context that determines the uses of language tests; (2) the nature of the language abilities we want to measure, and (3) the nature of measurement" (*ibid*.).

In terms of establishing usefulness, in a later publication Bachman and Palmer (1996, p. 18) set out a model for test usefulness as reproduced below:

Usefulness = reliability + construct validity + authenticity + interactiveness + impact + practicality

This model is, in fact, key for this thesis since this is in essence a usefulness study, however, I will not be considering each of the areas presented above in this study.

Firstly, impact as outlined by Bachman and Palmer (1996) concerns the impact of an assessment at two levels "a micro level, in terms of the individuals who are affected by the particular test use, and a macro level in terms of the educational system or society" (pp. 29 – 30). In the context of this study, however, as discussed in section 3.2 on washback, I will be considering the impact slightly differently, and for me this area comes into authenticity of task as well as validity, as will become clear in the following discussions.

Further practicality is another area I will not be considering since I am investigating existing assessments and quite simply, were they not practical then they would not be used. Additionally, interactiveness will not be included. Bachman and Palmer (1996) define interactiveness as "the extent and type of involvement of the test taker's individual characteristics in accomplishing a test task" (p. 25). Since the purpose of this study is to establish a practical model which can be applied by EAP practitioners within the time and resource constraints of their general workload, I believe that investigating this for each individual student would be impractical, thus rendering the model useless. Therefore, although I do not deny that the students' previous knowledge and experience will impact on how well they perform in an exam or assessment task, I do not believe that this needs to be considered in a usefulness study as proposed here. Thus, to summarise, in this thesis I will consider just three areas: authenticity, validity and reliability. The following subsections will now consider each of these areas in turn.

3.4.1 Authenticity

Watson Todd (2003, p. 153) highlights the fact that for the most part, EAP "promotes the use of authentic materials and tasks", with Malone and Montee (2014) explaining that "research on testing of second language acquisition has repeatedly demonstrated

that test-tasks should reflect real-life learning situations" (p. 3). Thus, what needs to be investigated is whether the test tasks being investigated have any level of authenticity, and, indeed, whether they really need to. This section will discuss authenticity and how it can influence, and be incorporated into, test design. I will also consider how to establish the level of authenticity.

Before considering authenticity, however, I come first to an important distinction made by Douglas (2010), based on Widdowson (1979), which is the difference between authentic and genuine language. Essentially, genuine implies that the materials that are being used to create tasks on which inferences about language ability will be made are genuine in nature (Douglas, 2010). For example, texts may come from a genuine academic article or listening tasks may come from a genuine recorded lecture. This is easier to achieve than authenticity, and although genuineness can contribute to authenticity, a genuine source does not automatically lead to an authentic task. Instead, what is required for an authentic task is authenticity in the uses made of the materials (Douglas, 2010).

So, how can test developers and users establish if a task is authentic? In order to do this, it is necessary to ascertain what the 'real world' language use situations actually are, which has been discussed in terms of academia by several authors. In a study carried out using a diagnostic language assessment procedure called Measuring Academic Skills of University Students (MASUS), Donohue and Erling (2012) explain that "much research into the assessment of EAP concentrates on the reliability, verifiability and validity of assessment procedures in the EAP context itself rather than on the relevance of what is assessed in EAP to what students have to write in the target language context" (p. 210). This is a point which I wish to take further to encompass all

of the target language use (TLU) tasks which students and lecturers identify in the focus groups so as to establish whether these tasks are being adequately reflected and assessed. In the case of this study, the TLU is any language use situation outside the English classroom, and could include situations such as lectures in the classroom or preparing for group work as well as reading for and producing written assignments and presentations. Donohue and Erling's (2012) study was, in their own words, "motivated by variations in academic attainment, represented by degree, module or assignment grade, between different groups at the Open University" (p. 211). Although very different research groups will be used in the present study, their research is considered a useful text for comparison of findings as well as offering an explanation of another potentially helpful evaluative model which can be integrated into the theoretical framework (MASUS).

Authenticity in a test situation, however, is something of a paradox, an issue which Douglas (2010) discusses, although he does not discuss it under the heading of authenticity. He highlights that it is "impossible to conceive of using language for no purpose at all, in no context at all. Yet, very often, that is what we are expected to do in language tests!" (p. 21). He (*ibid.*) goes on to explain that despite the contextual information or reasons provided in language test instructions, the test takers are still aware that the real purpose of the task they are carrying out is to demonstrate their language ability; and this is the paradox of authenticity. Test developers can make a test task realistic, with a complicated set-up and a great number of people involved, yet the purpose of the test is still essentially to test language ability; something which test-takers are well aware of (McNamara, 2000). So, no matter how authentic test developers make a test task, it is still only possible to make inferences about a student's ability in similar situations. Hence, although authenticity is a useful consideration when

evaluating the usefulness of a test, making a task truly authentic is not necessarily

practical, nor is it necessarily needed. Provided the task is adequate for inferring ability,

pure task authenticity is not required, although as Douglas (2010) highlights "if the test

purpose is to make inferences about a learner's language ability in some communicative

context, then the test should provide relevant contextual information" (p. 21). He

further explains that:

> there is a limit to how far we can take authenticity in language testing but [...] we
> must nevertheless make an effort to provide a context for language use in our tests
> to help ensure that the interpretations we make of test takers' performances will be
> valid.
>
> Douglas (2010, p. 26).

In terms of establishing what makes an authentic approach, the table outlined by

Bachman and Palmer (1996, pp. 49 – 50) and reproduced in Table 3 on the following

page. As can be seen, this table explains test task characteristics and is a key element of

the evaluative framework I am using. It is helpful because it changes slightly how

authenticity will be discussed in this thesis. The framework will be applied and used in

conjunction with descriptions of English for academic or specific purposes as an

approach to establishing what, exactly (if at all possible!) are the skills needed and how

proficiency in these can be demonstrated.

Table 3 - Test Task Characteristics: Bachman and Palmer (1996, pp. 49 – 50)

Task characteristics

Characteristics of the setting
Physical characteristics
Participants
Time of task

Characteristics of the test rubrics
Instructions
- Language (native, target)
- Channel (aural, visual)
- Specification of procedures and tasks

Structure
- Number of parts/tasks
- Salience of parts/tasks
- Sequence of parts/tasks
- Relative importance of parts/tasks
- Number of tasks/items per part

Time allotment
Scoring method
- Criteria for correctness
- Procedures for scoring the response
- Explicitness of criteria and procedures

Characteristics of the input
Format
- Channel (aural, visual)
- Form (language, non-language, both)
- Language (native, target, both)
- Length
- Type (item, prompt)
- Degree of speededness
- Vehicle ('live', 'reproduced', both)

Language of input
- Language characteristics
  - Organizational characteristics
    - Grammatical (vocabulary, syntax, phonology, graphology)
    - Textual (cohesion, rhetorical/conversational organization)
  - Pragmatic characteristics
    - Functional (ideational, manipulative, heuristic, imaginative)
    - Sociolinguistic (dialect/variety, register, naturalness, cultural references and figurative language)
  - Topical characteristics

Characteristics of the expected response
Format
- Channel (aural, visual)
- Form (language, non-language, both)
- Language (native, target, both)
- Length
- Type (selected, limited production, extended production)
- Degree of speededness

Language of expected response
- Language characteristics
  - Organizational characteristics
    - Grammatical (vocabulary, syntax, phonology, graphology)
    - Textual (cohesion, rhetorical/conversational organization)
  - Pragmatic characteristics
    - Functional (ideational, manipulative, heuristic, imaginative)
    - Sociolinguistic (dialect/variety, register, naturaltness, cultural references and figurative language)
  - Topical characteristics

Relationship between input and response
Reactivity (reciprocal, non-reciprocal, adaptive)
Scope of relationship (broad, narrow)
Directness of relationship (direct, indirect)

As Bachman and Palmer (1996, p. 45) have explained, "the key to designing tests that will be useful for their intended purposes is to include, in the test, tasks whose distinguishing characteristics correspond to those of the TLU tasks". This is, I believe, the key to the authenticity discussion, and a vital point for this study is the authenticity of the test task in relation to the TLU situation. Essentially, in order for an assessment to aid students with their academic acculturation, which was deemed a key element of EAP programmes in the literature review, the tasks included in it need to reflect the real-world situation; in other words, the tasks need to be authentic in style, or at least elicit authentic responses. This, however, is just the first area of consideration when establishing usefulness, and the following section will consider the validity of the assessment.

3.4.2 Validity

Validity is a very important area of consideration when attempting to establish the usefulness of a language test (Al Hajr, 2014; Aryadoust, 2013; Weir, 2005; Bachman and Palmer, 1996; McNamara, 1996; Bachman, 1990; Davies, 1990; Messick, 1989; Pilliner, 1968; Lado, 1961). Validity is essentially concerned with "the truth of the test, its relation to what it is intended to test" (Davies, 1990, p. 21), or as Lado (1961) explains "does a test measure what it purports to measure? If it does, it is valid" (p. 321). In their article on validity, Kane *et al.* (1999) briefly discuss validity in general, outlining that "neither an assessment procedure nor an assessment score is valid or invalid in itself. It is only when an interpretation is assigned to the scores that the question of validity arises" (p. 6). This is supported by O'Loughlin (2013) who states that "the validity of a test hinges critically on the interpretation of test scores and the uses to which they are directed" (p. 365). It is, therefore, the inferences made on the basis of the test performance that are valid.

When considering validity in general, McNamara (1996, pp. 15 – 16) explains that:

> the main issue is *how* and *how well* we can generalize from test performance to the criterion behaviour. *How* involves decisions of test *design*; *how well* is an empirical matter which can be investigated on the basis of *data* from trial or operational administrations of the test, supplemented if appropriate with additional data, for example, on the *predictive* power of the test.

This idea of validity inferences, that is the additional data, based on various different validity types is what is preferred by Aryadoust (2013). Before considering the different types of validity evidence, however, some discussion of the development of validity and validation studies in testing is required.

There are various different approaches to validation studies proposed in the literature, which have developed considerably over time. Brennan (2013) presents a brief overview of test validity/validation, which is a useful introduction to this section:

> the historical literature of validity/validation can be divided into three somewhat overlapping eras: (a) the beginnings to about 1950 in which prediction dominated along with a recognition of content validity; (b) about 1950 – 1990 in which discussions about construct validity predominated and (c) about 1980 to the present which witnessed the development of validation paradigms (most notably the argument-based approach) for addressing practical problems.
>
> Brennan (2013, p. 75)

It is clear from this, then, that there is a great deal of literature on which I could have drawn for this part of the study, and for the purposes of the discussions here, I have chosen to base my own validation discussions on the works of Bachman (1990) and Bachman and Palmer (1996). Before looking in detail at their work, however, it is necessary to consider the main theory which led to the development of their approaches - Messick (1989), whose seminal chapter in *Educational Measurement* is the basis for the models I am adapting for this study. For Messick (1989, p. 13), validity can be defined as:

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.

Messick (1989), like many other authors since he published his model (for example Lissitz and Samuelsson, 2007; Bachman and Palmer, 1996; Bachman, 1990) presents validity as a unitary, overarching concept. Despite preferring a unitary concept, he does see it as being faceted, explaining that "one facet is the source of justification of the testing, being based on appraisal of either evidence or consequence. The other facet is the function or outcome of the testing, being either interpretation or use" (Messick, 1989, p. 20). This model, he presented in a two-by-two matrix, reproduced below:

	TEST INTERPRETATION	TEST USE
EVIDENTIAL BASIS	Construct validity	Construct validity + Relevance / utility.
CONSEQUENTIAL BASIS	Value implications	Social consequences

Figure 1 Messick's Facets of Validity Framework (1989, p. 20)

One criticism of Messick's (1989) work has been that "the matrix was a mistake" (Shepherd, 1997, p. 6). Shepherd (1997) believes that this matrix (although not the rest of Messick's argument) over-simplifies the study of validity by implying that each of cells could be considered independently, with:

the traditional, scientific version of construct validity resid[ing] in the upper, left-hand cell and that consequences in the lower, right-hand cell are the business of moral philosophers and the politically correct.

Shepherd (1997, p. 6).

This criticism is reflected by Kane (2012) who explains that this model "does not provide a place to start, guidance on how to proceed, or criteria for gauging progress and deciding when to stop" (p. 8).

A further criticism of Messick's (1989) approach has been his inclusion of the consequences in validation study (McNamara, 2006; Popham, 1997). Starting with the work of Popham (1997), he contends that "the issues of social consequences [might] confuse the meaning of validity" (p. 9). He begins his criticism explaining that

> Although every right-thinking measurement person ought to be concerned about the consequences ensuing from a test's use, it does not follow that test-use consequences need to be linked to the … view that *validity* is rooted in the accuracy of inferences we derive from examinees' test performances. I believe, therefore, that the reification of consequential validity is apt to be counterproductive. It will deflect us from the clarity we need when judging tests *and* the consequences of test-use.
>
> Popham (1997, p. 9)

His argument against including consequences in a validity discussion centres around three main points: the accuracy of test-based inferences is more important in terms of building a validity argument; including social consequences in the validity discussion "will lead to confusion, not clarity" (Popham, 1997, p. 9); and "test-use consequences should be systematically addressed by those who develop and use tests, but not as an aspect of validity" (*ibid*.). As explained above, however, I am considering the consequences of test use in the area of washback, as well as in the area of test task authenticity as outlined in sections 3.2 and 3.4.1.

A final criticism of both Messick (1989) and subsequently of those who base their work on his model, including Bachman (1990) and Bachman and Palmer (1996), which I want to consider here is the preference for using construct validity as the overarching concept. I do not, however, believe this particular unitary, overarching concept to be

suitable for this study. Although as explained above, Messick's (1989) chapter is a

seminal work in the area of language testing, and one on which a number of authors

have based their own work, his preference for discussing construct validity has more

recently been considered "somewhat opaque in parts" (Weir, 2005, p. 21). Weir (2005,

p. 13) considers that validity is multifaceted, and that "different types of evidence are

needed to support any claims for the validity of scores on a test". He also explains that

these different types of validity "are not alternatives but complementary aspects of an

evidential basis for test interpretation". He adds "no single validity can be considered

superior to another" (*ibid.*); this may well be the case, but to establish the usefulness of

the assessments in question, examining all levels of validity is not feasible.

More recent work in the area of validation studies has considered different terminology

to be more useful than validity, or validity evidences. Kane (2006) for example

considers interpretive argument (IA) justification preferable to validity evidences or

construct validity. His reasons for doing so were explained in a later article, where he

highlights that although validity as a unitary concept was "conceptually elegant" (Kane,

2012, p. 7), it was "not very practical" (*ibid.*). In a later study, Kane actually updated

this approach with a change in terminology to interpretation/use argument (IUA) (Kane,

2013a, 2013b).

In discussing the IUA as presented by Kane (2013a), Brennan (2013) explains that it

was developed as a result of

> (a) the limitations of construct validity for providing a framework for addressing
> practical validation issues; (b) the reformulation of evaluation along lines
> consistent with validation, largely due to Cronbach; and (c) the efforts of [various
> authors] … to formulate validation paradigms that were faithful to the basic
> principles of validity qua theory.
>
> Brennan (2013, p. 76)

The strength of this approach, as Kane (2013a) explains, is that validation can be made more "manageable" (Newton, 2013) by setting out:

> the reasoning inherent in the proposed interpretations and uses of test scores by examining the evidence for and against the claims being made, including any evidence relevant to plausible alternate interpretations and uses.
>
> Kane (2013a, p. 8)

In doing this, Kane (2013a) claims that "the IUA provides an explicit statement of what is being claimed and thereby provides a framework for validation" (p. 8). Newton (2013), however, does question Kane's (2013a) approach since he believes that it is not clear whether there are "two *kinds* of argument, two arguments, or … just one" (p. 105, original emphasis), adding that "it is hard to find a clear illustration of a validity argument as distinct from an IUA" (Newton, 2013, p. 107).

There are, as have been presented above, many strengths to the IUA, not least of which is that it can be tailored to the validation of specific uses. Despite this, the focus of this thesis is to present a usefulness model which incorporates validity as just one area of investigation. Thus, since the IUA approach does not fully incorporate the other areas which I consider to be key to a usefulness study, namely authenticity and reliability. Hence, I do not consider IUA as being applicable to the present context.

Further, as highlighted by Brennan (2013) and Chahoub-Deville (2016) amongst others, the IUA approach is extremely complex and multifaceted. Although this is not beyond the skill or ability of EAP practitioners, it could be considered beyond their time constraints. Thus, since one of the purposes here is to present a practical usefulness model which can be applied with relative ease and within the limitations experienced by EAP practitioners as a result of their workload, the complex nature of the IUA approach to validity has been deemed inappropriate. That is not to say that should it be found that development is required in the form of action research, the IUA cannot be included in

that stage. In fact, as Newton (2013) and Brennan (2016) explain, the IUA is proposed

as a model for the development stage of assessment, thus although not wholly

applicable for the current study, it is an approach which could be used should

development and action research be deemed necessary.

In critiquing, and attempting to simplify Kane's (2013a) IUA, Sireci (2013) proposes a

three-step validation plan which is more accessible to practitioners and, hence, more

useful here than Kane. Sireci's (2013, p. 101) adapted IUA approach takes the

following form:

> (a) clear articulation of testing purposes, (b) considerations of potential test misuse,
> and (c) crossing test purposes and potential misuses with the five sources of
> validity evidence listed in the AREA *et al*. (1999) *Standards*.

Although a simplified approach to IUA, and one which is certainly more accessible to

EAP practitioners, it does not consider all of the areas of validity which I consider the

most relevant for the predictive validity discussion being proposed here.

To return, then, to the model being used as a basis for this study, that of Bachman and

Palmer (1996), Fulcher and Davidson (2007) highlight that their "notion of test

'usefulness' provides an alternative way of looking at validity, but has not been

extensively used in the language testing literature" (p. 15). Usefulness seems to me to

be a key term in evaluating assessments, thus it is the term I have chosen to focus on in

this study, and in doing so I will be extending the use of their theory, albeit in an

adapted form. The reasons for this have been partially discussed above, but before

going into more detail about my own approach, I first need to critique the work of

Bachman and Palmer (1996) a little more. To do this, I turn to Xi (2008) who explains

that although the usefulness framework is helpful in that it makes Messick's (1989)

framework accessible to practitioners, it lacks "logical mechanisms to prioritize the six

qualities and to evaluate overall test usefulness" (p. 179). A further criticism of

Bachman and Palmer (1996) from the point of view of this study is that their usefulness

model does not consider washback, nor does it address fairness (McNamara, 2006), or

ethical considerations (*ibid.*); all terms which have either been frequently used in

considering test evaluation, or have gained currency in more recent years. Despite these

criticisms, I still find the Bachman and Palmer (1996) model a useful starting point for

developing the framework I will be applying in this study, although I will not be

considering the whole model, nor will I necessarily be applying the concepts as outlined

in their book.

So far, then, I have established that the IUA as proposed by Kane (2013a) is not the

most suitable approach for the purpose of this study, and I have also established the

construct validity is not the most suitable overarching concept for use here, as well as

explaining why Bachman and Palmer's (1996) usefulness model need some adaptation

for use in this study. So, following the advice of Popham (1997), citing Messick

(1995), Weir (2005), Moss (2007) and Sireci (2007) I have chosen to examine the

assessments being investigated in this thesis for various types of validity evidence,

considering the construct, content, context and face validity of the assessments being

investigated.

That said, this does not mean that a unitary term, supported by various forms of

evidence, should not be applied. Rather, it is the term chosen which needs more careful

consideration. As an overarching concept for this study, I prefer to focus on the

predictive validity of the assessments, since a fundamental purpose of EAP assessment

is to predict how well students will perform in the situations which they will encounter

when they progress. Predictive validity is defined by McNamara (1996) as "the extent

to which the predictions we would make on the basis of test performance about subsequent performance in the criterion situation are likely to be borne out" (p. 21) and Al Hajr (2014) defines it as "the extent to which an assessment predicts future performance of assessed participants" (p. 122). Davies (1990) also discusses predictive validity, defining it as "the extent to which test results predict some future outcome" (p. 23).

For Aryadoust (2013), predictive validity "establishes that the observed scores represent the ability level of the test-taker relative to the challenges he or she will encounter in the target context" (p. 3). McNamara (1996) adds that "predictive validity is usually and often necessarily investigated after the test has become operational" (p. 22). Since the tests being investigated had already been in use for several years prior to this study, and are, for the most part, believed by the English team to be performance-based, communicative assessments aimed at eliciting performances which mirror real-life language use performances, predictive validity seems the most helpful overarching term for this study. In order to establish the predictive validity of the assessments, though, it is still necessary to consider various different types of validity evidence, which will be expanded below.

The first area for consideration is construct validity, a term preferred by Bachman and Palmer (1996, p. 21) who define it as

> the meaningfulness and appropriateness of the interpretations that we make on the basis of test scores when we interpret scores from language tests as indicators of the test takers' language ability, a crucial question is 'To what extent can we justify these interpretations?'

For Kane (2013a), construct validity is "the relationship between the test and a proposed interpretation" (p. 5). Construct validity, however, has had many different definitions, linked to quite different ideas. Weir (2005) explains that in 1980s American testing

literature, instead of considering whether a language test actually measured what it set out to measure, construct validity of a test was approached from the statistical perspective, as a way of validating "whether a test had measured a construct in individuals, which had a reality independent of constructs" (p. 17). He adds that in order to establish true construct validity, it is necessary to consider the construct that is being measured from the start, adding that

> The more we are able to describe the construct we are attempting to measure at an *a priori* stage the more meaningful might be the statistical procedures contributing to construct validation that can subsequently be applied to the results of the test.

> Weir (2005, p. 18)

Aryadoust (2013) considers construct validity as the most important element of any validity argument since "it establishes that the test's developers have determined the meaning, key elements, limits and boundaries of the construct targeted for assessment, and that the test indeed engages the target construct" (p. 3). This definition, similar to the one provided by Weir (2005) above, is supported by Al Hajr (2014) who explains that construct validity "indicates that an assessment instrument measures the skills and abilities (i.e. the constructs) that it is supposed to be measuring" (p. 122). Thus, in order to start developing a predictive validity argument, it is necessary to clearly define the construct that is being tested and this is, therefore, the starting point for examining the validity of the assessments being investigated.

In considering construct validity, a further area for investigation is the presence of construct-irrelevant factors, which Aryadoust (2013) believes can "confound measurement" (p. 2). This is, however, a complex area owing to the integrated nature of language. Aryadoust (2013) gives the example of a listening exam in which students may be unfairly penalised by being unable to read the questions. This, she believes, is not a fair test of listening skills, since reading questions is irrelevant to the construct

being tested – listening. In this case, the test could be considered not to be measuring the skills and abilities it is supposed to be (Al Hajr, 2014), yet it is almost impossible for testers to establish whether students have understood an audio message other than getting them to do something which could be considered construct irrelevant (i.e. write down the answer, carry out an instruction or respond verbally). Thus, all assessment tasks could be considered to have an element of construct irrelevance, therefore what needs to be investigated is whether the construct irrelevant factors unfairly penalise students.

The next area for consideration is content validity, which means that "an assessment is representative of the skills and content which [it] is supposed to measure" (Al Hajr, 2014, p. 122). In defining content validity, Kane (2013a) explains that it is an important form of validity evidence since "it is useful in evaluating the relationship between the sample of performance in the test and a larger performance domain that is of interest" (p. 5). A dated, although key text in testing literature, Pilliner (1968) explains that when attempting to establish the content validity of a test, "the test constructor must start with a clear conception of the aims and purposes of the area of learning to which [the] test is relevant" (p. 31). If the content of an assessment "enables the test user to evaluate the extent to which the testees have achieved the [...] aims [of the area of learning to which the test is relevant]; then the test possesses *content* validity" (Pilliner, 1968, p. 31, original emphasis). Essentially, for Pilliner (1968) content validity is concerned with the relationship between test or examination content and detailed curricular aims" (Pilliner, 1968, p. 32). Davies (1990) believes that "*Content* validity [...] is a professional judgement, that of the teacher or tester" (p. 23). Their views about this will, therefore, be invaluable to the current study.

Lado (1961) is a much quoted and referred-to source on language testing and his discussions of validity and reliability are of particular interest for this study. In terms of his thoughts on content validity, or "validity by content" as he terms it, he explains that by carrying out a content validity analysis of a test, it should be possible to establish whether or not the test tasks are adequately assessing the language elements under investigation, and whether there is an adequate sample of these being tested. If the language element that the tester wants to examine (for example pronunciation of a certain set of consonant clusters) is not adequately represented in the test tasks, then it cannot be considered valid (Lado, 1961). If the test is not valid, then it cannot be considered a useful test of language ability.

In contrast to the majority of authors writing about validity studies, Weir (2005, p. 19) believes that "the term *context* better accounts for the social dimension of language use". I would, however, argue that content and context validity are two different things. In Weir's (2005) study, context validity "is concerned with the extent to which the choice of tasks in a test is representative of the larger universe of tasks of which the test is assumed to be a sample" (p. 19); this definition is more closely linked to Bachman and Palmer's (1996) definition of authenticity, whereas content validity is linked to the communicative aspect of language use (McNamara, 2000).

Weir (2005, p. 20) goes on to explain that:

> …achieving context validity is not without its problems, given the difficulty we have in characterizing language proficiency with sufficient precision to ensure the validity of the representative sample we include in our tests.

He adds that "the greater the fit, the more confidence we may have in relating test performance to likely real-life behaviour" (Weir, 2005, p. 20) provided the test task in

question is considered to be valid when tested according to the other forms of validity as outlined above.

In terms of establishing both content and context validity, Ferris and Tagg (1996a, p. 52) present a table of pedagogical suggestions with suggested activities and tasks, which is reproduced below. This table offers a useful outline of the kinds of tasks which students would need to be assessed in order to ensure they are able to cope on their main courses. It is a useful tool for helping to establish the content validity and reliability of the assessments being investigated. With regard to content validity, Jordan (1997) explains that it is necessary to consider "the purposes of the test and then design… the test to reflect the areas to be assessed in appropriate portions" (p. 88). Thus, Ferris and Tagg's (1996a) table providing suggested pedagogical activities and tasks (PATs), reproduced below in Table 4, can be used to help with this.

Table 4 - Suggested Pedagogical Activities and Tasks (Ferris and Tagg, 1996a, p. 52)

Characteristic	Goal	Suggested activities and tasks
1. Class size	Make students aware of the characteristics of different class sizes.	Discuss strategies with students: interacting with teaching assistants (TAs), getting help from professors of large classes; discuss implications of very small classes; have professors or TAs visit to discuss these issues
2. Class participation	Expose students to various types of in-class activities; build their skills and confidence in classroom interactions.	Give students opportunities to take part in large-group and small-group discussions, ask and answer questions, agree/disagree appropriately with peers and instructors
3. Working with peers	Build students' abilities to work cooperatively both in and out of class.	Discuss pros and cons of collaborative work; assign in-class and out-of-class graded projects – library research, surveys, interviews, oral reports – that must be completed in pairs or groups
4. Oral presentations	Train students to make in-class oral presentations of different types.	Give students training and practice with a variety of speaking tasks: prepared formal speeches reporting the results of a small-group task or project in the class, brief oral summaries of course readings or case studies
5. Lecture comprehension	Expose students to different presentation styles and notetaking.	Model different lecture styles: invite other EAP and subject-matter instructors to guest-lecture; analyze audio- and videotaped lectures and transcripts for comprehension cues, strategies for asking and clarifying questions, and so on
6. Interacting with native-speaking peers	Help students develop confidence in working with peers who are not ESL students.	Set up conversation partner groups, social events, etc.; invite native-speaking students in for small-group discussions, interviews and so on
7. Interacting with professors	Help students feel more comfortable communicating with their instructors.	Invite subject-matter instructors to discuss expectations about office hours, telephone calls; discuss appropriate behaviour (e.g., addressing instructors)

The final area for investigation in terms of validity is face validity. Lado (1961) defines this as establishing test validity by "simple inspection" (p. 321). He (*ibid.*) believes that face validity is somewhat misleading, however, and is not a useful test of validity as there are many outside considerations which can impact on how well a student performs in a test, hence face validity is not going to be used for evaluating the language assessments under consideration. Although Lado (1961, p. 322) emphasises that "since language is a highly structured and highly complex activity we will need more than face validity to be convinced that at test measures the language element or skill it purports to measure", it is still important when considering usefulness since a task which 'looks right' can be one which reflects the pedagogical activities and tasks as outlined by Ferris and Tagg (1996a), and thus aids academic acculturation.

In defining face validity, Al Hajr (2014) explains it "signifies that an assessment looks suitable for its purposes" (p. 122). Further, Jordan (1997, p. 88) describes it as:

> the degree to which teachers and students think the test measures what it claims to measure. In other words, a subjective judgement of the test: does it seem reasonable, include appropriate content, have a suitable level of difficulty, etc. i.e. do people have faith in it?

Essentially, face validity "concerns the appeal of the test to the lay judgement", (Davies, 1990, p. 23). In the case of this study, the lay judgement will be that of the students, who will be asked to comment on the tests they were assessed with, and the subject-specialist teaching staff who will be given access to the assessments and asked to critique them during a focus group discussion.

As a way of summarising the above terms, Davies (1990) explains that validity is a way of overcoming "doubts about the 'mere' arithmetic of a test result" (p. 25). When considering the validity of the assessments, and ultimately establishing their usefulness,

the predictive validity of the tasks will be investigated, taking into consideration the construct validity, content validity, context validity and face validity of the tasks. If the tasks are considered to have high levels of *each type* of validity, then they can be considered to have sufficient predictive validity to make them useful.

3.4.3 Reliability

The final area for examination in the present study is the reliability of the assessment grades. Put most simply, according to Davies (1990), reliability is concerned with "the consistency of test judgements and results" (p. 21). Davies (1990) outlines four ways of establishing test reliability: "parallel forms (or versions) of the test; test-retest; split-half; and internal consistency" (p. 22). To further extend these terms, with parallel forms of the tests, two alternative forms of the same test are "given to an appropriate single group of learners with a short period between the two administrations" (Weir and Wu, 2006, p. 169). Heaton (1975) defines a parallel test of reliability involves "administering parallel forms of the test to the same group" (p. 156), with Hughes (2003) explaining that the results of both versions are then compared. This is done in order to "examine the equivalence of scores obtained from alternative forms of [the] test" (Bachman, 1990, p. 183). In contrast, test-retest reliability involves giving the test to the same group of students more than once (Hughes, 2003; Bachman, 1990; Heaton, 1975). The split-half test method is defined by Fulcher (2010) and Bachman (1990) as being when the test is divided into two halves which test equivalent skills in the same way, and Hughes (2003) explains that each half of the test must be truly equivalent "through the careful matching of items" (p. 40). In order to establish the reliability, each half of the test is then scored individually and the scores are then compared (Hughes, 2003). Finally, the internal consistency method is explained by Bachman

(1990) as being a way to establish "how consistent test takers' performances on the different parts of the test are with each other" (p. 172).

What is clear from the research relating to these approaches to establishing reliability is that these measures involve statistical measures of reliability (Fulcher, 2010; Weir, 2005). Further, all of these approaches are best used during the test development stages. Since the aim of this thesis is to develop a framework which could be applied to in-use assessments, and which can be applied by teaching staff who realistically would not have the time or resources to develop parallel versions of assessments, or carry out test-retest approaches, statistical measures of reliability will not be used for this study. Additionally, although some of the assessments used for this project are examinations with right or wrong answers, there is a great deal of subjective marking using generic criteria; something which could not be reliably compared statistically. In fact, Brown and Knight (1994) suggest that when it comes to criteria referenced tests such as those being investigated in this thesis, it is necessary to redefine reliability. They explain that statistical measures of reliability were "developed in psychometrics where a number of assumptions about the purposes of assessment and the nature of the data were made" (p. 19). In their opinion, such assumptions "do not hold good with criteria-referenced assessments, where the question is quite simply whether the best evidence is that the students has demonstrated the required competence" (Brown and Knight, 1994, p. 19). Therefore, a more descriptive approach to establishing reliability will be taken, in which grading criteria will be investigated carefully to establish how well they operationalise the skills and abilities deemed necessary.

Suskie (2009) explains that "grades alone may not tell us much about student learning, but the grading process can yield a wealth of valuable information" (p. 3), a good

starting point for discussing reliability using qualitative measures. Dunn *et al.* (2004), in fact, define reliability slightly differently from the statistical approaches listed above, explaining that it is "the extent to which other assessors would reach the same conclusions" (p. 17). For Weir, reliability is actually considered one of the various types of validity evidence. He (*ibid.*) calls it "scoring validity" (p. 23) and explains that it:

> concerns the extent to which test results are *stable over time*, *consistent in terms of the content sampling* and *free from bias*. In other words, it accounts for the degree to which examination marks are free from errors of measurement and therefore the extent to which they can be depended on for making decisions about the candidate (original emphasis).
>
> <div align="right">Weir (2005, p. 23)</div>

This is important in that it links directly to Bachman and Palmer's (1996) discussion of reliability. In discussing reliability, Douglas (2010, p. 3) uses the analogy of a rubber ruler. This analogy is particularly appropriate for discussing the reliability of a language test if we consider the fact that on the scale of proficiency, for example, the differences between the different levels of proficiency cannot be considered equidistant (Douglas, 2010, p. 4). Moreover, again as Douglas (2010, p. 4) emphasises, "on a language proficiency scale there is no true zero point". Hence it is not possible to say that one student knows three times as much language as another looking at just the test scores. The numbers alone, then, are not enough, and Douglas (2010) explains that a language test "might not be a very good measuring device" (p. 4) for assessing language proficiency, yet language tests continue to be big business. So, what can we do to ensure greater reliability from this 'rubber ruler' device?

Reliability is *"often defined as consistency of measurement. A reliable test score will be consistent across different characteristics of the testing situation"* (Bachman and Palmer, 1996, p. 19). However, as explained above, for this study it is not a

mathematical construct. Rather, in order to investigate reliability, I will be referring to the grading criteria and mark schemes that are used for the assessments under investigation, as advocated by Aryadoust (2013), McNamara (1996) and Brown and Knight (1994). What actually needs to be investigated is the way that the assessments are graded; what scales are being used?; how are they being used?; and most importantly, are the raters biased in any way? To answer these questions, "it is necessary to investigate and control for the effect of rater variation and scale characteristics" (McNamara, 1996, p. 3). What I will do when applying the framework is to

> tak[e] the test's scoring rubrics and procedures as evidence [so as to ensure] that the test's scoring processes have been standardized to a sufficient degree; that test taker behavior is consistently observed; and that scores are consistently awarded
>
> Aryadoust (2013, p. 3)

To do this, my starting point is to investigate the scale characteristics of the grading criteria, which will be done by considering various authors' ideas. The first of these are Bachman and Palmer (1996) who present a useful table (reproduced on the following page in Table 5) which can be used as a starting point for establishing whether the grading criteria are considering the most appropriate skills and grading them in the most appropriate way.

In establishing an evaluative framework for investigating the grading criteria being used for the assessments under investigation, I also refer to Donohue and Erling (2012) who offer some useful points for consideration, including "how to develop diagnostic descriptors which are discriminatory enough to pinpoint areas of strength and weakness in language use" (p. 211) as well as highlighting that it is necessary to consider "how to address the interrelating, multicomponential nature of language use" (*ibid*.). Further

points for consideration will come for the MASUS instrument[5], since the detailed sub-categories as presented in Donohue and Erling (2012) and Bonanno and Jones (2007) can be used as points to compare to the grading criteria being used for the writing and speaking assessments in question.

Table 5 - Areas of Language Knowledge (Bachman and Palmer. 1996, p. 68)

Organizational knowledge
(how utterances or sentences and texts are organized)

Grammatical knowledge
(how individual utterances or sentences are organized)
Knowledge of vocabulary
Knowledge of syntax
Knowledge of phonology/graphology

Textual knowledge
(how utterances or sentences are organized to form texts)
Knowledge of cohesion
Knowledge of rhetorical or conversational organization

Pragmatic knowledge
(how utterances or sentences and texts are related to the communicative goals of the language user and do the features of the language use setting)

Functional knowledge
(how utterances or sentences and texts are related to the communicative goals of language users)
Knowledge of ideational functions
Knowledge of manipulative functions
Knowledge of heuristic functions
Knowledge of imaginative functions

Sociolinguistic knowledge
(how utterances or sentences and texts are related to features of the language use setting)
Knowledge of dialects/varieties
Knowledge of registers
Knowledge of natural or idiomatic expressions
Knowledge of cultural references and figures of speech

MASUS is a procedure which "draws on systemic functional linguistics and genre-based pedagogy" (Donohue and Erling, 2012, p. 212). It was developed by Bonanno and Jones (2007) to assess students' "ability to think about a given body of knowledge in an analytical and critical way, and [their] ability to manipulate the resources of language in order to depict this thinking" (p. 2). MASUS is an evaluative instrument

---

[5] MASUS was first designed at the University of Sydney and is used to measure students' academic literacy. It stands for Measuring Academic Skills of University Students.

which has five assessment categories (Bonanno and Jones, 2007), the detailed sub-criteria of which were a helpful basis for the discussions with the subject specialists during their focus groups. Donohue and Erling (2012) explain that "rather than being a set of categories which can be transferred from context to context, the MASUS instrument is a set of principles drawn from systemic functional linguistic theory" (p. 215). This is of particular relevance to this study since EAP is a specific academic genre. When applying MASUS, Donohue and Erling (2012) explain that it is necessary to consider "how the instrument is developed and used in order to discriminate more effectively among the features of discourse in the new target language situation" (p. 215). It evaluates students in several areas: use of source material, structure and development of answer, academic writing style, grammatical correctness and qualities of text presentation (Bonanno and Jones, 2007). Although the MASUS instrument was "originally designed for a diagnostic procedure under … constrained conditions, where students write a single assessment text" (Donohue and Erling, 2012, p. 212), by using it as a point of comparison it will also aid the discussion of the findings in the present study.

Douglas (2001) and later Donohue and Erling (2012) believe that "subject knowledge is an integral part of specific purposes language assessment" (Donohue and Erling, 2012, p. 213). This also links in with Bachman and Palmer's (1996) discussion of considering the TLU setting. By using the criteria and sub-categories presented in the MASUS instrument, the texts produced by students can be seen as

> performing functions in context which are realised at whole text level (genre), paragraph level (discourse semantics), and sentence and word choice level (register/lexicogrammar), and by interactions within and between these levels.
>
> Donohue and Erling (2012, p. 212).

These categories can then be applied to the grading criteria to establish whether there is a link between the skills identified as key by the focus group participants.

3.5 Summary

The discussion in this chapter outlines the complexities of testing English for academic purposes, and has proposed some key areas for discussing assessment usefulness. As with the previous chapter, several generalisations have arisen from the discussion, which will warrant further investigation in this study:

1. The majority of discussion related to EAP assessment focuses on large-scale, high-stakes tests. Discussion of assessments developed in-house appears limited in nature.

2. Most sources related to test design are related to the initial design of tests rather than evaluating and reviewing tests which are already in existence. Proposing a model for this specific purpose fills a gap in the literature.

3. There is little advice or training available for in-service professionals who work on pre-sessional EAP assessment development and updating. Owing to the surge in the number of students on such programmes, this is a gap which warrants attention.

To summarise this chapter, there are three key areas which will be investigated in this study in order to establish whether or not the assessments currently in use at the case study institution are actually useful assessments of the students' language abilities for academic purposes. Taking the ideas of Davies (1990) establishing the usefulness of a test is about "*being explicit* as to what is being tested and what has been learnt and *controlling uncertainty* through statistical operations" (p. 53). This being explicit relates to test validity, and controlling uncertainty is actually ensuring test reliability; hence these two terms, as outlined by Bachman and Palmer (1996), are key when

evaluating test usefulness. In fact, as Davies (1990, p. 53) explains, validity and

reliability are "the chief themes of any testing discussion", but to these I believe must be

added the authenticity of the tasks and language produced, ensuring that they are truly

representative of the institutional pedagogical activities and tasks, that is the TLU

situations.

3.6 Language test usefulness evaluative framework

Taking into account the different approaches to validity and validation study outlined

here, and despite the criticisms of the model, I still find the work of Bachman and

Palmer (1996) a good starting point, especially since their main concern, like mine, is

one of building a usefulness discussion which is based on various forms of evidence.

Fulcher and Davidson (2007), in fact, highlight that Bachman and Palmer's (1996)

"notion of test 'usefulness' provides an alternative way of looking at validity, but has

not been extensively used in the language testing literature" (p. 15). For these reasons, I

have chosen to apply the adapted version of their model as displayed below in this

thesis. However, as discussed above, I will not be applying their concepts exactly as

they set out. The model below (Figure 2) provides a visual representation of the

approach I will be taking in the discussion section of this thesis.
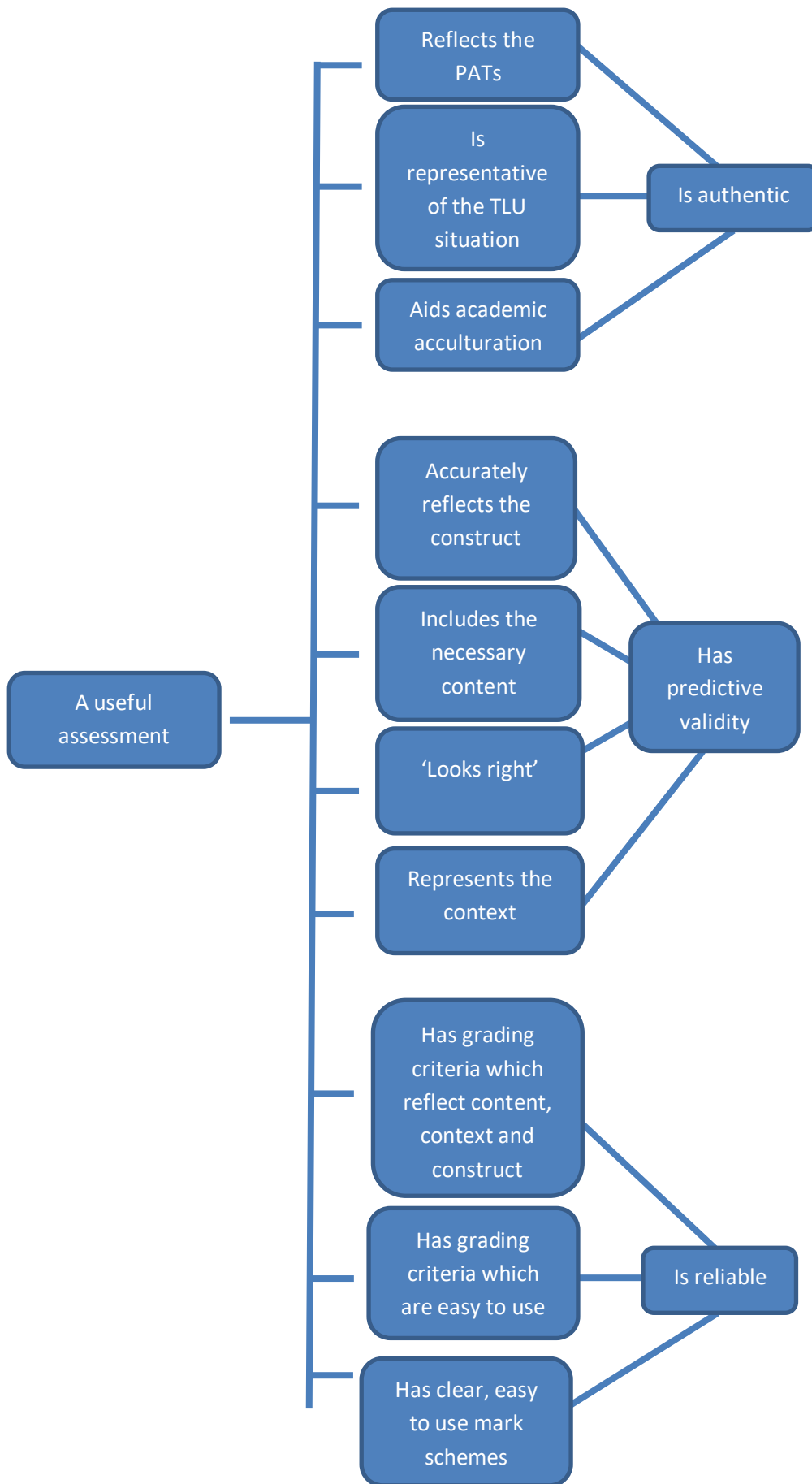
Figure 2 – Language Test Usefulness Evaluative Framework

To explain this model further, Table 3 below outlines how the key ideas discussed

above will be used to inform the analysis in this thesis:

Table 6 – Descriptive Framework for Establishing EAP Assessment Usefulness

Criterion	Application
Identifying the pedagogical activities and tasks (Donohue and Erling, 2012; Watson Todd, 2003; Douglas, 2001; Ferris and Tagg, 1996a, 1996b)	By establishing, through primary research, the key pedagogical situations in which the students will find themselves when they move onto their main courses of study, it will be possible to ascertain whether the EAP assessments reflect these, and whether they offer adequate assessment of the skills students will need to be able to cope in these situations. The skills will be outlined partly using the MASUS categories (Watson Todd, 2003) as well as other ideas from Ostler (1980) and Johns (1981). In highlighting the key TLU situations, it will also be possible to establish the extent to which tasks are authentic and provide academic skills development opportunities and aid academic acculturation.
Authenticity and academic acculturation (Douglas, 2010; McNamara, 2000; Bachman and Palmer, 1996)	By using the PATs as a way of representing the TLU situations (Bachman and Palmer, 1996) in which the students will find themselves, it will be possible to establish the authenticity of the task types as well as the language encountered and produced by the students. Although true authenticity in a language assessment is a paradox, in order to make accurate inferences about students' abilities, it is necessary to attempt to make use of authentic 'style' language and authentic 'style' tasks as a way of establishing whether students will be able to cope in similar target language use situations in the future. Authenticity also aids students with their academic acculturation, a key outcome of EAP programmes.
Types of validity evidence (Al Hajir, 2014; Aryadoust, 2013; Moss 2007; Sireci, 2007; Weir, 2005; McNamara, 1996; Messick, 1995 and 1989; Lado, 1961) leading to predictive validity	Various types of validity evidence as outlined by several authors will be considered as a way of establishing whether the assessments can be considered valid. These will include (but are not limited to) face validity, content validity, context validity and predictive validity. By investigating the types of validity evidence for each of the assessments, it will be possible to establish the level of validity for each of the assessments. This will be carried out by comparing the contents of the assessments to comments made by focus group participants. Essentially, the purpose of this part of the framework is to establish whether the tests measure what they purport to measure (Lado, 1961), and to establish whether the tests are adequately assessing the language and skills highlighted as most necessary for students to cope on HE level courses.
Reliability (Aryadoust, 2013; Douglas, 2010; Weir, 2005; Bachman and Palmer, 1996; Davies, 1990; Pilliner, 1968)	This aspect will be investigated by considering the reliability of the mark schemes and grading criteria rather than by considering the statistical reliability of the assessments when used across large student cohorts over a period of time, or when grades are compared to other assessments of a similar level. Reliability will consider how well the PATs and TLU situations as well as the language functions are described in grading criteria and mark schemes, and the levels of ambiguity and subjectivity in these.

In presenting this model and subsequent descriptive framework, I have now achieved

the third research objective, that is, to develop an evaluative framework for establishing

the usefulness of in-use EAP assessments.

CHAPTER 4

METHODOLOGY

4.1 Introduction

Having now achieved the first three objectives of this study, this chapter is the first step towards achieving the final four, and in it I will explain how I collected and analysed the data for this study. Before offering some critical discussion of the approaches I have taken in this study, however, the first area to cover is that of where I am coming from as a researcher. In order to do this, in the first section I will offer comment on the philosophical underpinning of my own position, since my own philosophical beliefs are the starting point for defending my choice of research approaches. In this first section, I will also comment on my role within the institution. This will be followed by a discussion of my approach to the collection and analysis of the primary data including a section on the ethical considerations related to this study; following this I will consider the choice of documentary data for this study and offer some justification for their inclusion. The final section will outline how I ensured the trustworthiness and credibility of the data. As with previous sections of this thesis, this methodology will be structured as answers to a series of questions: Where am I coming from?; Why choose a case study?; Why and how to use focus groups?; Who was asked?; Why involve multiple stakeholders?; What were the ethical considerations? Which documents were investigated and why?; How were the data analysed?; and finally: How were trustworthiness and credibility of the data ensured? Essentially, this chapter will explain how the data collected were used to expand on the framework presented in Figure 2 above.

4.2 Where am I coming from?

Wellington *et al.* (2005, p. 21) highlight that "it is impossible to take the researcher out of any type of research or any stage of the research process". They (*ibid.*) also explain that outlining one's position "can help counter charges of bias and partisanship". Sikes (2004) states that positionality is "where the researcher is coming from in terms of their philosophical position" (p. 18), and Wellington *et al.* (2005) add that it is the biography of the researcher, covering "how and where they are socially positioned, the consequent perspectives they hold and the assumptions which inform the sense they make of the world" (p. 21) which make up a person's positionality.

Positionality is discussed by various authors in terms of its influence on research methods (see for example Lichtman 2006; Denzin and Lincoln, 2003; Herod, 1999; Mullings, 1999). For this study, I take Martin and Van Gunten's (2002) definition, since they explain that:

> positionality is a concept that acknowledges that we are all raced, classed, and gendered and that these identities are relational, complex, and fluid positions rather than essential qualities
>
> Martin and Van Gunten (2002, p. 46)

Interrogating one's own position, then, is an important starting point for any discussion of methods and methodology since, as the researcher, in qualitative research at least, is "the primary instrument of data collection and analysis" (Lichtman, 2006, p. 12), they, therefore, decide exactly which data should be gathered and how best to gather them. Further, in gathering these data "all information is filtered through the researcher's eyes and ears. It is influenced by his or her experience, knowledge, skill, and background" (Lichtman, 2006, p. 12). Moreover, as Wellington *et al.* (2005) explain "the methodologies and methods selected will be influenced by a variety of factors, including: the personal predilections, interests and disciplinary background of the

researcher" (p. 99). This is echoed by Clough and Nutbrown (2012) who explain that "the positionality of the researcher affects research designs and processes as well as the ethical practices which are inevitably present throughout any study involving human beings" (p. 10). Further, Rossman and Rallis (2003, p. 284) highlight that "analysis and interpretation are an expression of [the researcher's] individual understanding of the phenomenon [they] study".

Important as it is, however, this first section will not outline my life history in detail; this has already been done in a previous paper (Clifford, 2012) and its inclusion here would not add anything to this study. What the remainder of this section will do, based on Wellington *et al.*'s (2005) advice, is firstly to explain my role within in the institution, and then attempt to outline my ontological and epistemological assumptions. These will then be drawn on throughout the remainder of this methodology to help justify the approaches that were used in this thesis.

In terms of my role in the institution, I have been a permanent full-time lecturer in EAP and modern foreign languages (MFL) at MUC since 2003. Although a long-standing member of staff, I am by no means senior within the English development centre, and for the first ten years at the university, my teaching hours were split equally between EAP and MFL teaching. It was only in 2013 that my teaching became exclusively EAP. As part of my role in the university, I was programme manager for the IEP for two years and for the International Foundation Diploma for seven years, as well as module leading on several of the IEP modules and the IFD English modules. As part of these roles, I have had responsibility for curriculum design and development and have been involved in writing, marking and moderating the EAP assessments, including those being investigated here.

Clough and Nutbrown (2012, p. 10) expound the inherently positional nature of social research such as the current study, explaining that:

> social research does not take place in isolation; people drive the research, they identify the emerging issues to be studied and they create – in context – the methods by which situations are further understood, and they communicate its outcomes to chosen audiences.

So, where am I coming from, then? As an educational researcher, I am approaching the study with a broadly constructivist position and with the epistemological assumption that, for the most part, knowledge is "experiential, personal and subjective and socially constructed" (Wellington *et al.*, 2005, p. 101). In discussing the constructivist paradigm, Denzin and Lincoln (2008) explain that it "assumes a relativist ontology (there are multiple realities), a subjectivist epistemology (knower and respondent co-create understandings), and a naturalistic (in the natural world) set of methodological procedures" (p. 32). This summary best defines my own beliefs.

As regards the current study, I am also working as an insider-researcher, a position which brings with it a whole series of issues, both positive and negative. Herod (1999) explains a great deal of the work published on the process of conducting primary research on other people "has tended to assume that there exists a simple and clearly discernible dichotomy concerning the researcher's positionality – either the researcher is an "outsider" *or* an "insider"" (p. 320). Deutsch (2004) also touches on this issue, stating that "all human relations encapsulate multiple subject-positions that take on different salience and meaning in various contexts" (pp. 897 – 898). It is not quite so clear cut, however, since in this study, I am considered both an insider- and outsider-researcher depending on the participants.

However, the issue of whether a researcher is an "insider" or an "outsider" is not as straightforward as some authors would suggest (Bell, 2005). In fact a researcher's positionality

> can shift depending upon a number of considerations, in the process disrupting the supposedly stable dualism of the 'insider'/'outsider' and the privileged positionality of the 'insider' that much literature on … methodologies has assumed exists.
>
> Herod (1999, p. 320).

In this study, although an institutional and community insider, I was also considered to be coming from outside the community.

The research participants, as will be discussed later, included members of the English team, of which I am also part, hence, I am a community insider in this context (Pillow, 2003). My split role at the university, however, also means that I may not be seen as a true community insider owing to the fact I am not solely an EAP teacher, with my role being split between EAP and MFL. Further, in terms of this group, I was also less experienced than all of the participants since they have wider work experiences, from other universities and work overseas, and more English-teaching specific qualifications when compared to my general language teaching PGCE. Thus, despite being part of this group, I could be seen as being on the periphery.

In addition to the English team there, of course, two other groups of research participants who were involved: subject-specialist lecturers and international students. To these two groups I was, to a certain extent, an outsider. For the subject specialists, I was seen to be an outsider as I am not part of the subject specialist communities which exist within the institution; in fact, the English department is not considered an academic department at all, instead it falls under the broad umbrella of student services. Therefore, for some of the subject specialists it became quite clear during the focus

groups that I was not considered an academic member of staff. This was not necessarily problematic, however, because, as Herod (1999) explains, the "outsider" researcher does have an advantage in that they constantly question and often take fewer things for granted simply because they do not understand things in the same way that "insiders" do. It also meant that I was able to ask probing questions during the focus groups which members of the 'academic' teaching team might not have felt comfortable asking (for further discussion of this, see below).

For the students, I was also an outsider in that I am a teacher rather than a student. That is not to say that being considered an outsider by these participants was a weaker position than being an insider; nor was the reverse true. Herod (1999) highlights that although "being perceived as an "outsider" has often been thought of as undesirable in the researcher process" (p. 322), this specific positionality in the eyes of the participants gave me the opportunity to ask probing questions, this time relating to the courses the students were currently studying or had recently completed and the skills they had felt it necessary to cope. This might not have been an option had the students perceived me as either an insider member of their group or a subject specialist since they may have expected me to know the answers to many of the questions I was posing.

In further discussing insider and outsider researchers, Herod (1999, p. 320) explains that it is often believed that the "insider" is "the most advantageous position in which to be since this gives the researcher a privileged position from which to understand … events". That is certainly the case here since I am researching my own institution and I therefore have 'insider' knowledge about systems and processes, particularly in my role as module leader as well as because of the curriculum design and assessment design work I have done. I am also in a privileged position in that I have access to documents

which are not readily available to those outside the institution. In addition, I also had the opportunity to involve staff and students from the university in the study; this is an option which institutional gatekeepers might not be happy to offer to those coming from outside MUC. Thus, I am starting from a position of privilege in terms of access to information.

That is not to say that being an insider is an exclusively privileged position; it also brings with it some disadvantages. One possible issue with my position as an insider was that it could have led to the participants, particularly the students, feeling reluctant to share honestly their thoughts and ideas (Franz, 2011). This could have been further exacerbated by the fact that I have taught some of the student participants, one of them for three semesters. A further issue of being an insider was that of potential bias. Indeed, I have a vested interest in the outcomes of this study: not only do I need the results to complete my thesis, but as a member of the team which developed the tests I am also interested in their efficacy. Additionally, my role as one of the test developers could have impacted on my evaluation of the assessments since demonstrating their efficacy is a key part of my job when attempting to defend the results they produce. So, in order to reduce the impact of this potential bias, I have attempted to be as open as possible with my own position from the start I have also attempted to ensure the credibility and trustworthiness of my data as is explained in more detail in part 4.6 below. I am not inferring here that my own position and interest will not have an impact on the results coming from this study, nor do I think that a purely objective approach is required, but I will attempt, by constantly questioning my results, to provide as balanced a view as possible when presenting the findings.

4.3 Why choose a case study?

A case study, as explained by Opie (2004) is "an in-depth study on interactions of a single instance in an enclosed system" (p. 74), adding that "its aim … is to provide a picture of a certain feature of social behaviour or activity in a particular setting and the factors influencing this situation" (*ibid.*). For Thomas (2009), the aim of a case study is "to gain a rich, detailed understanding of the case by examining aspects of it in detail" (p. 115). Rossman and Rallis (2003) define case studies as "in-depth and detailed explorations of single examples" (p. 104). Since the overall research question focuses on how EAP practitioners can establish the usefulness of in-use in-house EAP assessments, a case-study approach is the most suitable to take. This is not just owing to the ease of access to documentation and participants that I have as an insider-researcher, but also because of the approach I have advocated in the evaluative framework. As one of the aims of EAP programmes is academic acculturation, and one of the areas considered during the evaluation was the authenticity of tasks in terms of how well they reflect the PATs and TLU situations encountered by students, a single case-study approach is the most practical approach to take. Additionally, in applying this framework, it is likely that EAP practitioners will be working in one specific institutional context, therefore mirroring this in the application of my framework will offer a transferable methodology which can be used by others who wish to evaluate the usefulness of the assessments in use at their institutions.

To quote Giroux (1988) educational institutions "are places that represent forms of knowledge, language, practice, social relations and values" (p. 126). In short, an educational institution can be seen as a society with its own specific culture. Both Stake (2005) and Thomas (2009) discuss various reasons for using case studies. Stake (2005) in fact outlines three types of case study: intrinsic, instrumental and multiple case study.

Using Stake's (2005) and Thomas' (2009) definitions, the case study being used for this project is instrumental in that it is being "examined mainly to provide insight into an issue" (Stake, 2005, p. 445). To further quote Thomas (2009) "there is no intimation … that [I] will be generalising from [this] one case" (p. 115); instead the case is being studied because there are "particular circumstances surrounding the case that make it of special interest" (Thomas, 2009, p. 115). This is, thus, also an intrinsic case study since, as an employee at the institution and a key user and developer of the tests, I "have some kind of intrinsic interest in [the] case" (Thomas, 2009, p. 116). Thus, although not producing findings which are considered generlisable owing to the limited nature of case study research (Opie, 2004), the framework and approach to data analysis presented in this study could be considered transferrable and a useful tool for the evaluation of EAP assessment usefulness in a variety of HE settings.

To offer further comment on the generalisability, or not, of my findings, I turn first to Dzakiria (2012) who comments that, in terms of the generalisability of case-study research findings, for the researcher, "the aim of the research is to analyse a situation in order to understand it better and then to disseminate this new understanding for others to learn [from] (p. 44). In terms of the educational practitioner, "the aim of the research is to make use of fresh insights in effecting changes to his or her own context" (Dzakiria, 2012, p.44). That is my aim here. Kreuger (1993) in fact explains that the purpose of case-study research is not generalisability, rather, it is transferability.

Essentially, as explained above, I am proposing a transferable methodology and offering "closed generalisations" (Bassey, 1981, p. 84) which can be used to "stimulate … thinking about possible lines of action and … alert … to possible consequences" (*ibid*.). Further, as well as providing the advantages as explained above, a case study approach

also sits well with my own positionality. Scott and Morrison (2007, p. 17) explain that a case study is ideal for collecting data which are "qualitative and interpretive", as well as being in line with a social constructionist approach, thus it is an ideal choice for a researcher with my own ontological and epistemological beliefs.

4.4 Collecting the primary data

4.4.1 Why and how were focus groups used?

A focus group is "a particular type of group interview where the moderator (or researcher/evaluator) asks a set of targeted questions designed to elicit *collective* views about a specific topic" (Ryan *et al.*, 2013, p. 2 – original emphasis). According to Rossman and Rallis (2003), the aim of a focus group is "to encourage discussion and the expression of differing opinions and points of view". Bell (2005) is in general agreement with this, and her definition of the purpose of focus groups is that:

> participants will interact with each other …[and] perhaps … reach consensus about some aspect of the topic … and give a good airing to the issues which seem to be interesting or important to them.
>
> Bell (2005, p. 162)

This was the purpose in this study: to give participants the opportunity to discuss key areas relating to EAP assessment from their specific point of view, and to hopefully reach a consensus about them.

In terms of type of focus group, Ryan *et al.* (2013) offer various descriptions, one of which is the hybrid approach which generates a mix of personal opinion and collective experiences. This best describes the approach being taken in this study since participants have both personal and collective experiences. Personal, in that they all experience the impact of the EAP assessments on a personal level; collective in that they are part of one specific group (that is students, subject specialists or English teachers).

An advantage of the focus group design is that it can be used to elicit the views of (amongst others) key stakeholders and to yield rich description (Ryan *et al.*, 2013). As Rossman and Rallis (2003) explain, the focus group approach "assumes that an individual's attitudes and beliefs do not form in a vacuum" (p. 193), thus making them a suitable approach for me as a broadly social constructionist researcher. In fact, Ryan *et al.* (2013) emphasise the advantages, and indeed the appropriateness, of using focus groups in a social constructionist study such as this one, explaining that "from a social constructionist perspective, opinions are 'socially shared knowledge' or tacit knowledge that is generated, maintained, and changed through social participation" (p. 4). So, why am *I* using a focus group approach? I will answer this question in the remainder of this section.

Firstly, Linville *et al.* (2003) explain that using a focus group can "honor the knowledge and experience of people who typically do not have a voice" (p. 219). When looking at the research of Christison and Krahnke (1986), they explain that "curriculum design in ESL programs for academic preparation has, in general, failed to use the experience of students themselves as a basis for planning and decision-making" (p. 61). Although there is now considerably more account taken of student opinion in terms of syllabus design and course effectiveness, their voice is still lacking in terms of assessment usefulness. Thus, the use of student focus groups in this study was a way to ensure that this under-represented group has their voice heard. This point will be discussed further in the section below on why it was considered necessary to involve multiple stakeholders in the study.

In attempting to give a voice to students, as well as other stakeholders, in this study, the focus group was considered ideal since Morgan and Kreuger (1993) who highlight that this approach is ideal "when you need a friendly research method that is respectful and not condescending to your target audience" (p. 18). Since I involved my former students in the project, this "friendly" approach was key to encouraging them to discuss issues in a second language, with their teacher, and to do so honestly. Additionally, the fact that the focus group approach is "not condescending" made it appropriate to use with both subject specialists and the English team, meaning a similar approach could be taken with all research participants.

As outlined in the section on positionality, my ontological and epistemological beliefs mean that, for the most part, any study I undertake would be best done using "naturalistic" methodological procedures (Denzin and Lincoln, 2008). Moreover, Wellington *et al.* (2005, p. 100) explain that:

> if you take a social constructivist position it will be necessary to collect subjective accounts and perceptions that explain how the world is experienced and constructed by the people who live in it.

Essentially, as a social constructionist researcher I am investigating my place of work as an educational 'society' and so a focus group was a useful method of data collection since, as Colucci (2007) explains "humans, as 'social' beings have long been gathering together and discussing important issues in groups"; hence focus groups seemed to be an almost obvious choice. However, to avoid the issue highlighted by Colucci (2007) of the focus groups becoming simple group interviews, I followed her advice and included activities and pre-reading tasks to prompt and also focus the discussion.

With regard to the administration of the focus groups, I decided to moderate them myself. Franz (2011, p. 1382) explains that one advantage of this is that I had:

> studied the background information on the questions being asked, …explored the content driving the questions, and … pilot tested the questions with a group similar to those being studied.

Morgan and Kreuger (1993) also touch on this point and explain that in some studies there is a "need for a moderator who has a detailed familiarity with either the project goals or the participants' points of view" (p. 5). Since the aim of this project was to gain an in-depth understanding of the skills and activities students need in order to establish the test usefulness, gaining an in-depth knowledge of these PATs from the start of the research process helped me guide the participants of the focus group in the discussion of these areas. A further benefit of administering all of the focus groups myself was that I was able to ensure that the same questions, or at least the same topics, were addressed in all of the focus groups (Smithson, 2000).

The focus groups all took place on university premises as it was a mutually convenient location and meant that no participants incurred any extra costs. They took place during the work day to add to the convenience for all participants. Each of the groups was carried out in a classroom at the University, a location chosen for several reasons: firstly, to ensure that there was sufficient space for the first activity to be carried out and also to ensure that there was a table large enough for all participants to sit around so as to facilitate discussion. Additionally, following the advice of Kreuger (1993), the classrooms were "neutral and easy to find, a place where participants feel comfortable and relaxed" (p. 68). They also offered a quiet, private space allowing for uninterrupted discussions. The choice of a classroom is not without its potential problems in terms of power relations[6], however, as Elwood and Martin (2000) explain. They (*ibid.*) comment that a particular location chosen for an interview "situate[s] a participant with respect to other actors and his or her own multiple identities and roles, affecting

---

[6] See section 4.4.4 for further discussion of power relations.

information that is communicated … as well as power dynamics" (p. 652). With this in mind, the layout of the room was of particular importance. Hence, before the focus groups began, I set out the rooms to ensure that the participants and myself were sitting in a circle around a table. With the student groups, this was key to demonstrating that I was part of their discussion group, rather than just leading it.

The groups lasted between one and two hours depending on the group size, as was recommended by Ryan *et al.* (2013), with the longest discussion being with the English teaching team. Student group discussions were generally shorter than groups with teaching staff, mainly owing to the size of the groups.

Initially, the reason for using focus groups with the student participants in particular was for the 'safety in numbers' effect. This is supported by Sim (1998) who explains that in focus groups "participants may feel supported and empowered by a sense of group membership and cohesiveness" (p. 346). This was, however, not the case in reality since both student focus groups consisted of just two students owing to problems with recruiting student participants to the study.

In terms of the procedure of the focus groups, the schedules along with the questions used for each group can be found in Appendices E to G. With each focus group, I used a type of free-listing (Colucci, 2007) in which I asked each of the participants at the start of the discussion to list on sticky notes all of the language and academic skills and activities which they believed necessary for students to cope with HE level study. This was followed by a "pile sorting" activity (Colucci, 2007) in which participants were asked to discuss as a group and organise their notes in categories of their collective choosing. Throughout this part of the focus groups, I did little more than provide paper,

pens and notes to facilitate the activity, and occasionally answer questions regarding what was required. The discussions which occurred during these opening sections yielded some interesting data, and the results of the activities in the form of sticky notes categorised on sheets of paper, were photographed and subsequently used as a starting point for the thematic analysis of the data[7]. As was found by Colucci (2007), using activities rather than continual researcher questioning to prompt the discussion led to research participants commenting that they had found participating in the focus groups was a positive experience. Ryan *et al.* (2013) in fact outline that "a variety of stimulus materials … can be used for focus group facilitation" (p. 2). This is supported by Scott and Morrison (2007) who state that focus groups "usually involve collective activity(ies) to provide the framework for the interaction, reflecting the advice of Colucci (2007) from her article "Focus Groups Can Be Fun". Thus, in addition to the free-listing and pile-sorting activities, the subject specialist participants were also asked to review the assessments and grading criteria currently in use on the English programmes between 2010 and 2014 prior to attending the focus groups. This made it possible to discuss these participants' opinions regarding the reliability, validity and task authenticity. This was not necessary for the EAP team or the student focus groups since these stakeholders already had experience of the assessments. The focus groups, therefore, followed the structure below.

---

[7] See Appendix H for examples of these photographs.

Stage 1
- Individual freelisting of the skills and activities needed to cope with HE level studies at MUC

Stage 2
- Group pile sorting of freelisitng activities into themes or key areas

Stage 3
- Discussion of the pile sorting outcomes

Stage 4
- Discussions of the EAP assessments and grading criteria in use at MUC in relation to the first three stages
- Comments on and recommendations for improvement for the assessments in relation to the discussions

Figure 3: How the focus groups were carried out

In planning a focus group, Ryan *et al.* (2013) state that it is important to consider the type of evidence required from the focus group when deciding on its design, and that was also the case here. The aim here was to establish the participants' views on the skills and attributes required of students if they are to be able to cope with HE level study, something which needed to be considered from the participants' points of view rather than imposed by me as a researcher. I also needed to establish the institutional culture as experienced by the students, and imposed by the subject specialist staff, rather than how it is perceived by the English team, myself included. This evidence is best gained from speaking to these people.

For the EAP teachers, then, the aim was to offer the group a rare opportunity to evaluate their own work, discussing the strengths and weaknesses and interrogating the decisions on which assessment choices were made. For the subject specialists, the purpose of the focus groups was slightly different in that the aim was to gather collective shared experiences about the institution's academic culture, as well as the factors required to cope with HE level study, what should be tested, and how it should be tested whilst also eliciting their opinions about the current assessments.

In the case of the student participants, the purpose of carrying out the focus groups was two-fold. Firstly, the discussions provided additional details about the institutional culture. Secondly, the focus groups offered students an opportunity to comment on the quality of the assessments on which their progression decisions had been made, and to give their opinions regarding the various validity types outlined in the previous chapter. Special emphasis in the student focus groups was given to validity, with discussion prompted around the face validity and content validity of the assessments, as well as how well the assessments reflected their experiences on their main programmes of study.

4.4.2 Who was asked?

Having outlined how the focus groups were administered, this section will deal with the theoretical approach taken to choosing the sample of participants to take part in this study. As a broadly social constructionist researcher, the methods being used to collect the data needed to consider the experiences of the people who make up the society I am researching (Wellington *et al.*, 2005, p. 102). In this case, as outlined above, the 'society' is MUC, and the research participants came from three key stakeholder groups within this society: the EAP team, subject specialist lecturing staff and students who had been tested with the tests under investigation.

The process of choosing and recruiting participants for the study was not an easy one. The easiest group to recruit was the EAP teaching team, although this was not without its problems. The whole team, both permanent and agency staff, was contacted and invited to participate. All members of the team were keen to take part, but owing to needing to fit around teaching and examination boards, the EAP focus group was

planned for the end of June 2014 to coincide with a quieter period at the institution. Unfortunately, this meant that, in the end, only four members of the team could participate in the group: three full-time, permanent members of staff, and one member of agency staff.

The subject specialist participants were recruited in a similar way in that they were invited by email and mutually convenient dates were chosen. In order to select a suitable group of participants, I considered the structure of the university carefully. MUC's programmes are split across four larger schools: Hospitality, Tourism and Events; Business and Marketing; Education and Community; and Sport and Creative Studies. The majority of international students take programmes in the school of Hospitality, Tourism and the school of Events and Business and Marketing, with fewer students entering courses in the other schools. Despite the smaller numbers of students on some programmes, however, I felt it necessary to invite participants from all schools as the English programmes aim to prepare students for entry onto any programme at any level and in any school. Additionally, since one of the aims of the EAP programme is to help students adjust to the new institutional culture, I needed to establish if there is a 'general' institutional culture for them to adapt to.

In total, 12 participants took part across two focus groups. These participants were representative of all university schools apart from Business and Marketing. Despite this, owing to the cross-institution nature of several of the modules taught at MUC, some of the participants did teach on business and marketing programmes, so all schools were to some extent represented. In the first focus group there were three participants from the school of Hospitality, Tourism and Events, two from the school of Education and Community and one from Sport and Creative Studies. The second group

had five participants with four from the school of Sport and Creative Studies and two from Hospitality, Tourism and Events.

The student participants were less straightforward to recruit. Over the space of six months, 212 students were invited to participate in the study. As with the other groups, all potential participants were sent an email invite with details of the study. Since very few responded, selection criteria were eventually limited to those students who responded to the requests to participate. Having had positive responses from seven students, two focus groups took place, however only four students participated: two in each group. Although initially students from all schools within the University were invited to participate, those who did were studying programmes in Hospitality, Tourism and Events and Creative Studies.

A key part of the selection criteria for the students was that they had been enrolled on either the SEP or the IEP before starting their main course of study. The student participants were in different stages of their course: one was in the second year of her degree, one was in the dissertation stage of her MSc programme, one had just started the final year of his degree and one was studying on a pre-masters diploma. Despite the small number of participants, then, a broad range of programme levels was reflected in the sample.

There were several reasons why I chose to continue with the focus groups despite the small numbers, not least of which was the fact that the students had agreed to give up their time to participate and had attended at the agreed time and place. I did not, therefore, feel it appropriate to cancel the focus groups on the day and rearrange; this could have impacted on my relationship with the participants and led to no information

being provided by the student participants. This last-minute change would also have had an impact on the informed consent that students gave; they had agreed to participate in focus groups, not individual interviews so changing the data collection approach would immediately have impacted on this.

Further, although I will be discussing the ethical considerations in more detail below in section 4.4.4, I feel it necessary to touch on ethics at this point owing to the fact that there were only two participants in the student focus groups. Additionally, considering the advice from Smith (1995) that all participants should be treated equally, it did not seem 'equal' to cancel the focus groups and instead expect the students to participate in individual interviews. It could also have been considered 'deception' (Denscombe, 2008). Further, not only could a change in the data collection have led to unequal treatment of participants, it could also have impacted on the power relations. As a person potentially seen as being in a position of authority (Karnieli-Miller *et al.*, 2009), the students may have been less open in an individual interview than they were in the focus groups; this is also discussed by Wilkinson (2011, p. 71) who explains that:

> compared with most traditional methods, including the one-to-one interview, focus groups inevitably reduce the researcher's power and control. Simply by virtue of the number of participants involved in the research interaction, the balance of power shifts away from the researcher.

This was true, even with only two participants in each discussion; there were still more students than researchers involved in the focus groups.

Finally, I decided to continue with the small student focus groups based on comments from Morgan (1996). As Morgan (1996) explains, there are some researchers who believe that focus groups "must meet some specified set of criteria, typically that they consist of structured discussions among 6 – 10 homogenous strangers in a formal

setting" (p. 131). Morgan (1996, p. 139), however, explains that this rather limiting

definition fails to take account for the reality of research and also highlights that

> what makes the discussion in focus groups more than the sum of separate
> individual interviews is the fact that the participants both query each other and
> explain themselves to each other … such information offers valuable data on the
> extent of consensus and diversity among participants. This ability to observe the
> extent and nature of interviewees' agreement and disagreement is a unique strength
> of focus groups.

With the purpose of the focus groups discussions being to allow participants to discuss

together to describe the key skills needed to cope with HE level study, as well as to

build a picture of MUC's institutional culture, it was felt at the time that these

discussions, even with only two participants, would be valuable. Moreover, as with the

staff focus groups, the free-listing and pile-sorting activities during which the

participants worked together were key to understanding the students' perceptions of

international culture, and the key activities, skills and language needed to cope with HE

level studies at MUC. By moving from focus group to interview, this key discussion

activity would have been lacking, and I would have missed a great deal of important

information. Thus, despite the small numbers participating, I believe that continuing

with the focus groups was the best decision.

In total, there were 20 participants in this study, broken down into the following: four

English teachers, 12 subject specialists and four students. As regards sampling theory,

Ryan *et al*. (2013, p. 2) explain that "focus group research often utilizes some type of

purposive sampling scheme" adding that focus group participants are "purposefully

selected based on a significant, homogenous characteristic" (p. 2). Franz (2011, p.

1381) explains that when choosing people to participate in a focus group, they:

> should be selected based on characteristics they have in common related to the
> purposes of the study or project. This may include homogeneity in occupation,
> social class, levels of education, or characteristics.

For the present study, the majority of participants were selected based on their availability. The homogeneity of the English team lay in the fact they were from the same department and had experience of using the assessments being investigated. In the case of the subject specialists, they were homogenous in that they were all subject specialists at the case-study institution; the students were homogenous in that they had all studied on one of the pre-sessional English programmes. For these two groups, however, that was where the similarities ended. Thomas (2009) explains that purposive sampling "involves simply the pursuit of the kind of person in whom the researcher is interested" (p. 104), and indeed the sample chosen for this study was not chosen because it was in any way representative of the make-up of the institution, but rather to get a broad view of opinions from across the different schools.

Further advice from Franz (2011) is that participants should "be unfamiliar with each other, since familiarity can inhibit disclosure" (p. 1381). This was not, however, possible in this case since the research was carried out at a small HEI where the majority of staff members are familiar with each other. Despite Franz's (2011) advice on this matter, though, I found that the familiarity between the staff members led to a very frank and honest discussion as group members felt comfortable with each other and were happy to share their thoughts. The students were in groups which did not know each other owing to timetable clashes and general availability, though, and I felt that this possibly detracted from their honesty since some of them were less confident speaking a second language in front of strangers. So, contrary to the advice of Franz (2011), I believe that some level of familiarity amongst the participants actually enhanced the discussion within the focus groups since participants were already comfortable with each other.

4.4.3 Why involve multiple stakeholders?

Having now outlined my own position, who was involved in the study and how they were recruited, I will now move on to a more in-depth defence of my choice of research participants, considering the theoretical framework and key definitions of EAP. As explained in Chapters 2 and 3, the evaluative framework for this study will focus on four key areas: the pedagogical activities and tasks that the students face once they move onto their main courses of study, the links between authenticity and acculturation, the predictive validity of the assessments and the reliability of the grading criteria and mark schemes.

As discussed in the previous section, three key stakeholder groups were included in the study: EAP lecturers, subject-specialist lecturers and former pre-sessional English students. Starting with the EAP lecturers, these were probably the most obvious participants, but their inclusion still needs some further justification. In terms of the evaluative framework, EAP staff input was invaluable for establishing the predictive validity of the assessments, in other words whether "the predictions we … make on the basis of test performance about subsequent performance in the criterion situation are likely to be borne out" (McNamara, 1996, p. 21). This requires expertise in the field of applied linguistics, a level of expertise which was only brought to the study by the EAP specialists. Focussing specifically on the content validity, that is "a professional judgement, that of the teacher or tester" (Davies, 1990, p. 23), clearly the EAP lecturers are able to comment directly on this. Davies (1990) adds that investigating content validity is a way of overcoming doubts about the sample of language being tested, and the only way to ensure this was, again, to consult with the EAP team.

As explained above, however, these tests are multifunctional in that they are also considered proficiency tests. A proficiency test, as explained previously, is used "to measure overall mastery of [English] or how well prepared [the student] is to use [English] in a particular setting" (Madsen, 1983, p. 9). A further definition of a proficiency test is provided by Douglas (2010, pp. 1-2) who explains that:

> proficiency tests are usually intended to help us make predictions about what learners will be able to do with the language in communicative situations outside the classroom.

In order to establish whether the assessments in question really are making accurate predictions, I turned again to the language teaching specialists since they are trained in this area and were thus best placed to be able to comment on this point.

To return to Tajino *et al.*'s (2005) definition presented in the previous chapter, EAP is concerned with ensuring the students have a level of practical English language skills which "meet the requirements of concerned people" (p. 35), speaking to these 'concerned people' about the usefulness of assessments of these language skills seems to be an obvious choice. Not asking these subject specialists, then, would appear to be a serious oversight, yet literature which includes their opinions on even the content of EAP courses is both limited and somewhat outdated.

Considering the evaluative framework in light of Tajino *et al.*'s (2005) comments above, subject specialists' opinions have been sought on all four points. It is their courses that the students are moving onto, so they are better placed than the EAP team to be able to describe the PATs that these students will encounter and to discuss the institutional culture. With regard to the predictive validity of the tasks, and linking quite closely to the idea of authenticity and face validity, the subject specialists have been asked about how the assessments look in relation to the assessments and tasks that

students will have to deal with on progression. Hence, their opinion on the face validity and context validity of the assessments is invaluable. In addition, these participants were asked to comment on whether they thought that the tasks were authentic in relation to the PATs which they outlined, and also whether they thought this was actually needed. In doing this, the subject specialists were also consulted as a way of establishing the content validity of the assessments. Essentially, their observations about the usefulness of the assessments in use at the time of study were helpful not only in deciding on the level of test task authenticity required, but also in deciding whether it was necessary to redesign the assessments as well as how to do this in terms of planning future research.

Further to these points, the grading criteria were also discussed with the subject specialist staff. Again, their discussion of the PATs came into play, and they were asked to appraise the grading criteria to consider how these activities and tasks are reflected within the criteria, and whether they thought these representations were both accurate (and thus reliable) or even necessary.

The final group included in the study were the students who had studied on the pre-sessional language programmes and then moved onto their main programmes. Christison and Krahnke (1986) explain that "curriculum design in ESL programs for academic preparation has, in general, failed to use the experience of students themselves as a basis for planning and decision making" (p. 61) and it can be assumed that this is also the case with assessment design and evaluation. More recent studies on curriculum design[8] have considered in part the opinions of students, yet their voice still appears to be lacking when it comes to evaluating the assessments used at the end of a course.

---

[8] See for example Ferris and Tagg (1996a) and Ferris and Tagg (1996b)

Additionally, it is these students who have been required to adapt, albeit only temporarily, to the new academic culture, so it is they who are best placed to comment on the usefulness of the assessments in terms of academic acculturation.

Not considering these users' opinions, therefore, would have led to a clear gap when it comes to the framework being used for this study. Although the PATs can be described by the subject specialist lecturers, the student opinions of what is important in this area need to be considered. It is the students who are faced with, and therefore having to cope with, these new sets of academic demands and so establishing whether these have been effectively reflected and tested in the end-of-course assessments needs to be considered from their point of view. This is also the case with the authenticity discussion. The EAP team have attempted to produce tasks which they believe mirror as closely as possible the tasks that students will encounter on their main course of study, yet it is still unclear whether this is really necessary. Although the subject specialists were able to discuss this point to some extent, the students' opinions were most valuable since they were the only group who could really comment on the need for authenticity. In order to establish whether authenticity is truly necessary for EAP assessment tasks, student opinions are key.

Thus, I want to investigate the experiences and personal experiences of representatives of the 'society' I am studying. In this case, the 'society' is the university where I work, so I am also bringing an insider perspective to the study (Herod, 1999). For the current study, the most important members of this 'society' are the EAP team, the subject specialists onto whose courses the students progress, and the students themselves. If I am to fully appreciate and understand the impact and quality, and thus the usefulness, of the assessments being investigated, then I need to do it by considering their points of

view. I feel that missing out one of these key stakeholders would lead to an indefensible gap in my research data. In addition to these reasons, including the views of multiple stakeholders provides a way of enhancing the credibility of the findings through triangulation as recommended by Graneheim and Lundman (2004).

4.4.4 What were the main ethical considerations?

Since my research was "probing human existence in detail" (Brinkmann and Kvale, 2005, p. 157), ethical considerations have been important throughout both the research and reporting of this thesis. As explained in the rationale and earlier in this section, I carried out this research at my place of work, which brings with it a number of potential ethical issues (Bell and Nutt, 2002; Shordike *et al.*, 2017). The first area of ethics for discussion is the process of ethical approval. Approval was sought from both the University of Sheffield (see appendix M for ethical approval letter) as well as senior management at MUC for the research to take place. Senior management at MUC approved the research project on the understanding that the institution was anonymised in the report findings, hence the institution being referred to as MUC throughout the discussion.

In terms of ethical procedures related to the participants, prior to participating in the study, on the advice of Denscombe (2008) and Miles and Huberman (1994), they were sent the participant information sheet (appendices A to D) as part of the recruitment process and asked to sign a consent form before the focus groups took place. They were then given a copy of the signed consent form and a hard copy of the relevant information sheet. Further, as stated in the participant information sheets, all data were anonymised in the discussion of the findings, with only participant initials used. This plus the anonymization of the institution was done to ensure that it is not possible for

readers of the study to identify the participants. The amount of information provided to participants was done so on the advice of Denscombe (2008) who explains that "[participants'] participation must always be voluntary, and they must have sufficient information about the research to arrive at a reasoned judgement about whether or not they want to participate" (p. 145). In providing the level of information for the participants that I did, my aim was to allow them to give this voluntary informed consent. Further, in providing the information sheet as part of the recruitment process, this gave all participants time and space to read this through before agreeing to participate.

With regards to anonymization of participants, this was not done without care consideration. As will be touched on as this section progresses, power relations are a key consideration in qualitative research (Miller and Bell, 2002; Bravo-Moreno, 2003; Karnieli-Miller *et al.*, 2009), and as a way to address some of the power relations issues, that is to say, as a way to re-balance the power, participants were asked how they should be represented in the research report. They were asked if they would prefer verbatim quotes or 'tidied up' versions of what they had said during the focus groups, and they were also asked how they wanted to be presented. For the latter question, I gave them three options: pseudonyms, their initials or complete depersonalisation in the form of a code, for example SS1P1 (subject specialist group one participant one). The decision from participants regarding quotes was unanimous; they all opted for verbatim quotes. In terms of how they should be represented, there was some disagreement, with two participants initially opting for pseudonyms, and the remainder preferring initials. My final decision was to choose using initials for since this was the majority preference, and on discussion with the two participants who preferred pseudonyms, the assurances that

the institution was anonymised and that they, therefore, would not be identifiable, resulted in their agreement to this approach.

As alluded to above, a further area related to ethics in relation to the participants is that of power relations (Elwood and Martin, 2000; Miller and Bell, 2002; ; Bravo-Moreno, 2003; Karnieli-Miller *et al.,* 2009; Baiardi *et al.*, 2015). Smith (2006) highlights that there are two broad conceptualisations of power: a structural view of power and a postructural view of power. She explains that the structural view sees power "as an inscribed capacity, something which is appropriated by particular individuals or organisations" (Smith, 2006, p. 644), adding that from this point of view "power is configured across society so that particular individuals 'possess' power which they can use to achieve certain outcomes" (*ibid.*).

To take this idea further, Karnieli- Miller *et al.* (2009) explain that, in fact, power relations shift during the different stages of the research process. They (*ibid.*) describe power as moving from being entirely in the hands of the researcher during the initial stage and participant recruitment, to the researcher appearing "entirely dependent on the participants' willingness to take part in the research and share their knowledge" (Karnieli-Miller *et al.*, 2009, p. 282). Having collected the data, the power returns to the researcher as they analyse the data and produce the report, but could be placed back into the hands of the participants during the validation process (Karnieli-Miller *et al.*, 2009). I will return to the validation process later in section 4.6. For this section, I will consider the power relations between researcher and participants during the focus groups.

As explained in the rationale, as well as in the first section of this chapter, I am an experienced, long-standing, although not senior permanent EAP and MFL lecturer at the institution. In terms of the English teaching team, then, as explained above I am to a certain extent a group insider (Herod, 1999). I have worked at the institution longer than any of the participants, although I was less experienced as an EAP lecturer than all of the participants in that group. Thus, in terms of power relations, I could have been seen to have been in a less powerful position than the participants, and I certainly did not hold any authority in terms of either my experience in EAP or my role in relation to this group. These two factors could have, and at times did, lead the participants to dominate the discussion during the focus group and at times to take it in a very different direction. It was only through some negotiation with the participants that I was able to bring the discussion back to the point. This happened in particular during a discussion which began about the skills needed in comparison to the assessments in use, but very quickly moved to discussion about the merits of various language learning apps. There was a definite imbalance of power during this focus group which was skewed in favour of the participants rather than me as the researcher/facilitator.

My position as a permanent member of staff raises the first issue relating to power relations and vulnerable participants (Denscombe, 2008) which may have impacted on the information provided by BW, who was employed by MUC through an agency. In an attempt to combat this possible issue, and on the advice of the University of Sheffield ethics team, it was made clear in the participant information sheet for agency staff (appendix D, paragraph four) that participation in the study would not impact on any future contracts with MUC. Further, this particular participant was included as part of the English team, and his potential difference in terms of position within in the institution was not commented on at any point during the focus group. He is a well-

known member of the team, and was included as part of the English team throughout. Further, his additional experience within the institution as someone who has supported students in classroom situations outside of the English department gave him additional knowledge not possessed by the rest of the group. In terms of the discussions about the classroom situation on main courses of study, then, BW had greater knowledge than the rest of the team.

In terms of the subject specialist staff, as explained above, it was possible that I was considered a group outsider. Certainly, in terms of the knowledge of institutional practices their roles as subject specialist lecturers means that they are in possession of considerably more knowledge than me as an English lecturer. They are in effect the authorities on institutional norms and conventions, and certainly in terms of power relations in the focus groups they were in a more powerful position when carrying out the freelisting and pile-sorting activities. Despite this, it could also be said that as the researcher of this project, I could also have been seen as being in a powerful position. My experience as an EAP lecturer, as well as my research in the area of EAP assessment prior to the focus groups, meant that I was considerably more knowledgeable about the topic being researched. This knowledge could have put me in a position of power too. Hence, during the focus groups with the subject specialists, the power relations were fluid with the dominance and authority moving from me as the facilitator with the knowledge about assessment and English practices, to the participants as the subject specialists with the knowledge I needed to be able to complete my research (Karnieli-Miller *et al., 2009).

Finally, there is the power relation between me and the student participants. As explained above, my position as a lecturer, and in fact as a former teacher of some of

117

my participants, I could be seen as being in a position of power in relation to the participants (Karnieli-Miller *et al.,* 2009). However, as Bravo-Moreno (2003) and later Brinkmann and Kvale (2005) have highlighted, even participants who could be considered to be in a vulnerable or powerless position, can exert considerable power during primary data collection by employing a variety of devices from evading questions to talking about different topics. These derailment tactics were not employed by the students, however. Further, as the researcher, I chose the topics of discussion and the approaches being used in the focus groups, thus giving me further power as explained by Baez (2002). However, despite some sources seeing them as being 'low-status' or 'powerless' participants, taking Karnieli-Miller *et al.*'s (2009) idea that it is the participants who hold the power in the data collection stage, the students were consulted as the experts in two areas: they had experience of the English programmes and the assessments used on them, and they had the experience of academic life at MUC outside the EAP classroom. Thus, they were uniquely placed to give the most useful information for the study; this, in effect, made them the most powerful participants since they were in a privileged position in terms of knowledge about the area of research (Smith, 2006).

The final ethical consideration was that of the results of the study itself. This study has been funded by my place of work, and is investigating the usefulness of assessments I had a hand in writing. This brings with it a whole raft of possible issues. Firstly, there is the situation of not wanting to make my sponsoring institution 'look bad'. In reporting unfavourable findings, I could present the institution in a bad light which might not sit well with the senior management. This issue was dealt with by anonymising the institution. The second issue is that the results of the study may have (and in fact did at times) demonstrate shortcomings in the assessments used during the

time of study (that is, 2010 – 2014). Ethically, this is something that could not be ignored, and, as was one of the purposes of the framework presented in the previous chapter, the study has in fact led to a complete overhaul of all assessments in use at MUC on pre-sessional English programmes based on the findings presented here.

4.4.5 How were the primary data analysed?

The primary qualitative data collected for this study are "considered to provide holistic accounts of institutions, individuals, or other phenomena" (Scott and Morrison, 2007, p. 33). In this case, one aim of the primary data collection was to provide an account of practices at MUC so as to establish the institutional culture and identify the most important PATs and TLU situations with a view to ascertaining the relative authenticity and predictive validity of the EAP assessment tasks. In order to organise these data into useful themes, the constant comparative method was applied. Constant comparison "involves going through [the] data again and again … comparing each element – phrase, sentence or paragraph – with all of the other elements" (Thomas, 2009, p. 198). For Corbin and Strauss (2008) constant comparison is a key element of thematic analysis, with the similarities and differences being clearly classified and "incidence found to be conceptually similar [being] grouped together under a higher level descriptive concept" (p. 73).

To look at the thematic analysis more closely, a key question relating to the analysis of focus group data and arising from the research of Ryan *et al.* (2013, p. 15) is "did the focus group participants establish common ground in conversation or primarily act as individuals?". This is a crucial question, and is one which I applied to the analysis of each of the focus groups. Both Kidd and Parshall (2000) and Sim (1998) mention that when analysing focus group data, a vital starting point is to find areas of consensus and

areas of dissent or controversy, and this was, indeed, my own starting point. O'Leary (2004) explains that the aim of thematic analysis is "to move from raw data to meaningful understanding" (p. 196), adding that this "is a process reliant on the generation/exploration of relevant themes" (*ibid.*). For Rossman and Rallis (2003) thematic analysis "typically emerges from the deep familiarity with the data that comes from categorising" (p. 282). For this study, although the overarching themes of pedagogical activities and tasks, reliability, validity and authenticity are guiding the evaluation of the assessment usefulness, an inductive analysis of the focus group data was still required. Having identified the key, overarching themes, the next area was to establish areas of consensus under these headings, and thus I was able to ascertain the institutional culture, and identify the most frequent and important PATs and TLU situations. In doing this, I could then evaluate the assessment tasks, considering how they reflected these, and whether they actually needed to[9].

The focus groups were recorded using audio and video-recording devices and following this, I transcribed them to produce a text which I could then subject to content analysis. In carrying out inductive analysis of the focus group transcripts, I discovered various themes, which I was then able to link back to the evaluative framework for the study. I was also able to draw on my own experiences both as a student and as an EAP practitioner. This inductive approach, combined with drawing on my own experiences is recommended by O'Leary (2004). Some of the themes used for analysis arose from an exploration of the repetition of key words, which was carried out by manually colour-coding the transcripts of each focus group. Following O'Leary's (2004) advice on this approach, I did this by systematically reading through the data to "find all instances of a particular word (or phrase), making note of its context/meaning" (p. 196).

---

[9] An example of how the themes were developed from the transcripts can be seen in Appendix I

In the case of this study, some of the themes were relatively easy to identify in that the participants were asked to categorise the skills and attributes they identified in their first task. Table 7 below presents the initial categories which the participants used to organise their data:

Table 7 – Themes provided by participants during the focus groups

English Development Centre	Subject Specialist Group One	Subject Specialist Group Two	Student Group One	Student Group Two
• Note-taking • Paraphrasing • Language skills • Speaking • Critical stuff • Life skills / study skills • Structure • Understanding assignments • Dig out / research skills	• Study skills • Oral skills • Academic – paraphrasing • Academic – referencing • Academic – critical thinking • Language – general skills • Language – objectivity and style • Language – words used • Language – basic language • Personal	No themes provided	• Academic writing • Speaking • Research skills • Listening • Useful skills	• Writing • Reading • How to do research • Getting used to local daily life

These categories were then used as the themes during the first round of coding. Further reading and categorisation of the data led to more themes being drawn out of the discussions had by participants around the categories they had themselves produced. That said, the second subject specialist group listed and grouped skills and attributes, however they did not provide any headings for their categories. Despite this, it was still possible to identify themes during their discussions. For the most part, I tried to avoid using too many themes generated from the literature and approaching the data with pre-determined categories in order to avoid the potential issue raised by O'Leary (2004) of

forcing the data to fit to my expectations and "not being able to see alternative explanations" (p. 197).

Having identified the key themes, my next step was to explore "the relationship between and among various themes" (O'Leary, 2004, p. 197). This was done by following O'Leary's (2004) advice and investigating the data to establish where "the use of certain words and/or concepts [was] correlated with the use of other words and/or concepts" (p. 197). I also looked at how themes correlated within and across the different groups. In approaching the thematic analysis, the advice of Rossman and Rallis (2003) was also helpful. In looking at theme generation, they explain that "thoughtful analysis demands a heightened awareness of the data, a focussed attention to those data, and an openness to the subtle, tacit undercurrents of social life" (*op. cit.* p. 282).

With regard to the deductive categories which arose from the literature review, these were used as "preliminary categories to focus data gathering" (Rossman and Rallis, 2003, p. 282). Rossman and Rallis (2003, p. 282) add that these deductively produced categories "are often expressed in [focus group] questions and guidelines". Vaismoradi *et al.* (2013) explain that deductive categories such as those arising from my literature review are useful for a study which is using thematic analysis. In fact, for me these categories were used to "provide direction for the data gathering" (Rossman and Rallis, 2003, p. 282) in that they helped me to produce a focus group schedule with key questions to guide the discussion.

Rossman and Rallis (2003, p. 282) explain that "themes … often emerge during intensive analysis as the researcher locates and describes more subtle processes". This

was the case with this study, in that initial themes arose from the focus group

categorisation activities, what Rossman and Rallis (2003) refer to as "indigenous

categories" (p. 282). Further themes were then drawn out of the data during the

constant comparison stage of the analysis, and were constructed by me. Rossman and

Rallis (2003) explain that these new categories "represent the etic view and are often

referred to as analyst-constructed categories" (p. 283). As was highlighted by Rossman

and Rallis (2003), these categories did not always correspond exactly with "the

categories of meaning used by the participants" (p. 283).

Having analysed and coded the data, the accounts from the key stakeholders have been

used to expand on the evaluative framework I presented at the end of Chapter 3.

Essentially, it is these data which have helped me to establish the following things:

- The institutional academic culture

- The most frequent TLU situations at MUC

- The most important PATs at MUC

In doing this, I was then able to evaluate the EAP assessments in use between 2010 and

2014 to establish whether they reflected these, or provided students with the opportunity

to develop the skills needed to cope with these.

4.4.6 How were the data reported?

Kreuger (1999) advocates the use of direct quotations when reporting focus group

findings. Considering this point, and also following the advice of Cordon and

Sainsbury (2006), the majority of data from the focus groups are presented in the

following chapters as a verbatim report of the participants' words. This verbatim

reporting approach was discussed with the participants both during and after the focus

groups. Verbatim quotations were chosen since one of the areas of evaluation for the assessments was to establish the extent to which they prepared students for their new academic culture. Direct quotations from the focus groups have, therefore, been used to illustrate typical institutional PATs and identify university TLUs, and as such present an authentic view of MUC's academic culture. This approach has also been used when considering their opinions of the assessments in use between 2010 and 2014.

In choosing to present the participants' words directly, where appropriate, rather than a 'tidied up' version of their words, I am also offering a voice to the participants, some of whom are often under-represented when it comes to research on EAP assessment (Dooey, 2010; Christison and Krahnke, 1986; Johns, 1981).

4.5 The documentary data

As well as collecting primary data in the form of focus group discussions, a large part of the evaluation for this thesis was based on evaluating documentary data. These data fell into two categories: the EAP assessments and the assessments used across all programmes at MUC. This subsection will consider which documents were included in the analysis, and discuss how I went about analysing them.

4.5.1 Which documents were included and why?

The documents chosen for analysis in this thesis fall into two broad types: the EAP assessments and the assessments used on all other programmes at MUC. The EAP assessments being investigated are those which were used at the end of the 10-week and 6-week pre-sessional summer English programme and 13-week and 26-week Intensive English programmes between 2010 and 2014. They cover all key skill areas and students encounter various academic tasks including researching for and writing essays

and reports, academic reading skills, note-taking in lectures and participating in debates and seminars. These assessments have been chosen because they have, over, the years, been designed with a view to assessing the skills the English team believe the students need in order to cope with their main course of study.

Choosing appropriate EAP assessments to investigate for this study was problematic because, as is clear from the literature review, language assessments can have many functions and can take many forms (Bachman and Palmer, 1996; Davies, 1990). Having considered all the assessments in use at MUC, the final decision was to investigate the assessments, mark schemes and grading criteria used on the pre-sessional Intensive English Programme and Summer English Programme. The rationale behind this was that these courses focus purely on language teaching, and have the aim of preparing students for entry onto their next course of study. The assessments, therefore, are multipurpose in nature in that they are achievement, i.e. they measure progress and are a way of establishing what has been learned of the course syllabus (Douglas 2010; Madsen, 1983; Davies, 1977; Davies, 1968) whilst also being a way to establish students' language proficiency in that they are also used to "make predictions about what learners will be able to do with the language in communicative situations outside the [English language] classroom" (Douglas, 2010, pp. 1-2). The usefulness of these assessments, therefore, merits investigation.

Further to the complex nature of the assessments, the tests used on these programmes are also high-stakes in nature because, to put it at its simplest, the grades are the only factors on which final progression decisions are made. Essentially, if students do not achieve the 60% pass grade (equivalent to B2 on the CEFR) on these assessments, their academic career could be at an end, or at least postponed until they have reached the

level of English deemed sufficient to be given a place on their next programme of study. Ensuring that the assessments are testing and developing the right language and skills in the most appropriate ways, then, is key.

The decision to investigate all other assessments used at MUC was much simpler. It is these which form a significant part of the PATs which students are required to carry out at MUC, and it is these which dictate a large number of the TLU situations in which students will find themselves when they progress. With this in mind, not including these assessments in the discussion of these areas would be a clear oversight. This follows the advice of Scott and Morrison (2007) who explain that one reason for using documents in research is "to contribute to the development of key concepts and the constructions of research instruments" (p. 75). The documents used in this study were used to identify the key concepts, that is the PATs and TLU situations in the case-study institution.

4.5.2 How were the documents analysed?

The documents included in this study were analysed in quite different ways. Starting with the assessments encountered by students on their main programmes of study after the IEP or SEP, simple content analysis was used. Rossman and Rallis (2003) explain that content analysis is "an overall approach, a method and an analytic strategy … [which] entails the systematic examination of [text] to objectively document patterns" (p. 198). Although an approach which does not necessarily fit with my own ontology and epistemology, a quantitative review of the types of assessment and key assessment terminology was carried out. The results of this assessment audit were then used to provide basic descriptive statistics of the 980 individual assessments used across all programmes at MUC in the academic year 2013 - 2014. In doing this I established the

most frequently used task types and task words used across the University. This helped to build a picture of the institutional academic culture as well as providing an insight into the key PATs and TLU situations.

Coming to how the EAP assessments were evaluated, these were analysed using part of the theoretical framework outlined at the end of the Chapter 3. Initially, the authenticity of the different assessments was considered in relation to the PATs and TLU situations identified by the focus group participants. The purpose here was to establish whether the assessments were providing students with the opportunities to develop key study skills.

Having done this, the assessments were then investigated for predictive validity, considering their face validity in relation to the PATs and TLU situations and their context validity in terms of the academic context at MUC. Additionally, how well they represented the construct (listening, reading, writing or speaking) and their content validity was considered in relation to the comments from all three stakeholder groups.

Finally, the mark schemes and grading criteria were investigated for their reliability. In doing this, I considered how well grading criteria represented or described the constructs being tested, as well as how effectively they operationalised the key skills and attributes required for coping with academic life at MUC. Additionally, I considered the potential for subjective marking and the level of detail in the exam mark schemes. In analysing the EAP assessments in this way, I implemented the evaluative framework, enhanced by comments from the stakeholder groups and details from the general assessment audit.

4.6 How were validity and reliability of the data ensured?

With validity and reliability being such important terms for investigating the assessments being used in this study, it is also important that I apply such terms to the results of my own research. Starting with validity, this term has been discussed in relation to qualitative research by many authors (for example O'Donoghue, 2007; Scott and Morrison, 2007; Kreuger, 1993; Lincoln and Guba, 1985). O'Donoghue (2007) explains that in qualitative research, validity "may be considered under two headings … internal validity and external validity" (p. 196).

Elo *et al.* (2014) present a useful checklist for researchers attempting to improve the trustworthiness of a content analysis study such as the one being presented here. Their recommendations focus around the three key phases of qualitative content analysis, namely the preparation phase, the organization phase and the presentation phase, and the rest of this section will follow these three phases.

4.6.1 The preparation phase

Elo *et al.*'s (2014) first series of questions relate to preparing for carrying out the data collection. They focus on three areas in this phase: the data collection method, the sampling strategy and selecting the unit of analysis. The first two areas, data collection and sampling, have already been detailed in sections 4.4.1 to 4.4.3 above. To briefly recap, I chose to use focus groups to collect data from a purposive sample of participants from three key stakeholder groups, namely students, the English teaching team and subject specialist lecturers at a case study institution.

In discussing validity in relation to this phase of the study, Kreuger (1998) also comments that pilot testing any questions being used is necessary "to ensure that they

were understood" (p. 68). In order to pilot the approach being used for this study, I organised three pilot study focus groups with representatives from each of the stakeholder groups who were willing to participate but were unable to take part in the main focus groups. The results of these focus groups were positive, in particular the free-listing and pile sorting activities were found enjoyable. Despite the general positive outcomes, the student participants recommended that some of the questions be slightly re-worded to aid comprehension. Further, the English teacher participants recommended setting a time limit for the free-listing task. It was as a result of the subject specialist focus group, and the subsequent discussion with those who took part, that I chose to video record the focus groups. It was also as a result of this group that I chose to use three recording devices to audio record the focus groups, since on checking the recording I had made during this group, I discovered that the device I was using had failed, a common pitfall which can impact on reliability as identified by Easton *et al,* (2000). Finally, it was on the advice of both the subject specialist and English team participants that I re-wrote my explanation of the free-listing and pile sorting activities to ensure clarity.

Further areas for consideration in this phase arise from Easton *et al.* (2000) whose article on avoiding common pitfalls in qualitative data analysis has been particularly useful. Their article relates in particular to practical issues which could impact on the dependability of data.

Firstly, as explained above, the possibility of equipment failure can have a significant impact on the dependability of data, so I chose to use three audio recording devices as well as video recording equipment to ensure I captured all of the data. In addition, and again on the advice of Easton *et al* (2000), the classrooms chosen for the focus groups

were chosen paying attention to their location in the building so as to reduce the potential problem of "extraneous noise" (Easton *et al.*, 2000, p. 705). A final dependability issue highlighted by Easton *et al.* (2000) is that of transcription errors. To avoid this, they recommend that "ideally, the researcher should also be the interviewer and the transcriber" (Easton *et al.*, 2000, p. 707), an approach I took in this study.

The transcription and subsequent analysis of the data brings me to the third area, selecting the unit of analysis, which has been discussed in section 4.4.4. In outlining the methods chosen for analysing my data, as well as in presenting the information in section 4.4, then, I have gone some way to beginning to establish the trustworthiness of my data in that I have explained the methods used to collect the data, choose the participants and analysed the data. This approach, as recommended by Denscombe (2008), O'Donoghue (2007) and Kreuger (1993), is my attempt to openly outline my approach in this methodology section as a form of audit trail, demonstrating how the research was carried out and how the findings were analysed.

4.6.2 The organisation phase

Elo *et al.* (2014) offer various methods to ensure the trustworthiness of data during the next stage, which is where the data analysis is considered in more detail and concerns areas such as "whether the categories are well-created, what the level of interpretation is and how to check the trustworthiness of the analysis" (Elo *et al.,* 2014, p. 5).

These three areas can be discussed together, although I will start with the creation of the categories. The approach taken to this is outlined in section 4.4.4 where I have explained that the categories arose from participants themselves., however, in order to establish that these were truly representative of what the participants had said, I used a

member-checking approach to the transcripts and the emerging categories. This approach is recommended by Denscombe (2008), Graneheim and Lundman (2004), O'Donoghue (2007) and Kreuger (1993). Although not without its problems, as highlighted by Forbat and Henderson (2005), I chose to ask the research participants to verify my transcripts of the focus groups and my representations of the categories they elaborated during the free-listing and subsequent pile sorting activities. In doing this, I was aiming to establish if "a recognisable reality ha[d] been produced in the view of participants" (O'Donnoghue, 2007, p. 197). Despite the fact that Lincoln and Guba (1985) explain that there is no way of knowing for certain that I have completely got it right, this verification by the participants was one step in establishing, if not the validity, then at least the credibility of the data (Denscombe, 2008; Guba and Lincoln, 1985).

An additional area for consideration here is how to establish the overall representativeness of the data (Elo *et al.,* 2014). This relates to the two terms validity and credibility. O'Donoghue (2007) explains that reliability "often refers to the capacity for the study's findings to be replicated" (p. 197). This premise is, however, is based on the assumption that there is a single, objective reality that can be observed, known and measured and is a problematic concept when applied to a study such as this one. I prefer, therefore, to use Lincoln and Guba's (1985) term dependability, which they define as establishing "means for taking into account both factors of instability and factors of phenomenal or design induced changes" (p. 299).

In an attempt to enhance the validity, or at least the credibility of the findings, I also followed Lincoln and Guba's (1985) advice of triangulating my data. Denscombe (2008) defines triangulation as using "contrasting data sources to bolster confidence that the data are 'on the right lines'" (p. 297). This was done firstly by involving the three

stakeholder groups and comparing and contrasting their comments in response to my questions. This approach is also recommended by Graneheim and Lundman (2004) who explain that "choosing participants with various experiences increases the possibility of shedding light on the research question from a variety of aspects" (p. 109) is a key way of establishing credibility in qualitative data. Triangulation was also achieved through comparing and contrasting primary and secondary findings to establish areas of agreement and disagreement. Finally, in comparing these two data sources to the assessment audit, I was able to establish not only the key PATs related to life outside the EAP classroom at MUC, I was also able to establish how this compared to academic life in the West in general. In doing this, I was able to compare and contrast three sets of findings with the assessments under investigation to establish their usefulness.

4.6.3 The reporting phase

The final phase, that of reporting the data, involves ensuring that results are reported "systematically and carefully, with particular attention paid to how connections between the data and results are reported" (Elo *et al.,* 2014, p. 6). In order to achieve this, I have split the findings into two chapters, structured around the evaluative framework presented at the end of Chapter 3.

A further approach to ensuring the credibility of my data in the reporting phase was through the use of quotes in reporting the findings. Graneheim and Lundman (2004) recommend that "show[ing] representative quotations from the transcribed text" (p. 110) provides a way to demonstrate differences both between and within the participant groups. With this in mind, then, I chose to use verbatim quotes for the most part when

presenting the data in the findings, thus reducing the impact of any interpretation which I may make of the data when paraphrasing.

In following the steps set out by Elo *et al*. (2014), my aim has been to ensure that the data presented here are done so transparently, with the research process outlined as clearly as possible, and the analysis procedures explained in detail. Thus, although the findings of this study cannot be expected to be generalisable owing to the case-study approach being taken, the findings presented here can be seen to be a faithful representation of the situation in the case-study institution.

4.7 Summary

To conclude this section, then, this is a mainly qualitative case-study which is drawing on a variety of data sources in order to triangulate the findings and establish the usefulness of the assessments in use at MUC between 2010 and 2014. By bringing in the thoughts and opinions of the three stakeholder groups, as well as the university-wide assessment audit, I was able to add detail to the evaluative framework presented on page 61, which was then used to evaluate the usefulness of the in-house pre-sessional EAP assessments. As discussed in the rationale, my aim with this study is not to provide generalizable results, but rather to offer a transferrable methodology which can be applied by other EAP practitioners to establish the usefulness of in-house developed assessments. The results of this evaluation can then be used to inform decision-making in terms of changes to be made to assessments, or even carrying out action research.

Having now achieved objectives four and five of this study, that is to collect stakeholder opinions regarding the language and skills needed to cope with academic life at MUC and to carry out an institution-wide assessment audit, the next part of this thesis will use

these, along with the assessment audit data, to discuss the assessments in use. The first discussion chapter will consider the pedagogical activities and tasks, the target language use situations and the academic culture at MUC. Following this, Chapter 6 will discuss the predictive validity and reliability of the assessments.

PART II: FINDINGS AND ANALYSIS – IMPLEMENTING THE

FRAMEWORK AND ESTABLISHING TEST USEFULNESS

Introduction

So far in this thesis I have achieved the first five objectives, providing a short history of EAP, critically evaluating the literature relating to language assessments, assessment evaluation and academic acculturation in the literature review. In Chapter 3, I presented the evaluative framework for establishing EAP assessment usefulness, which culminated in the model presented on page 85. The methodology chapter then explained how I achieved research objectives four and five, that is how I collected stakeholder opinions and carried out the assessment audit. This second part of the thesis will relate to objectives six and seven, and in the following two chapters I will demonstrate how the stakeholder opinions and details from the assessment audit can be used to establish the institutional culture, TLU situations and most important PATs. These findings will then be used to identify the authenticity, predictive validity and reliability of the assessments and their corresponding grading criteria and mark schemes, ultimately establishing the usefulness of the assessments.

Phakiti *et al.* (2013) explain that English language proficiency, although critical to international students' academic performance, as will be discussed in Chapter 6, is not the sole determining factor. As explained in the literature review, Dooey (2010) found that international students often see their pre-sessional programmes as "a sort of training ground for future challenges" (p. 192), therefore Chapter 5 will use the findings from the assessment audit and the focus groups to establish the pedagogical activities and tasks, the target language use situations and the institutional culture into which students will be entering once they start their main programmes of study. I will then consider how well the assessments in use at the time of study reflected these activities and tasks and represented the institutional culture, also evaluating how well the grading criteria and mark schemes used for the assessments oprationalised these. This chapter will also

consider the face validity of the EAP assessments. Chapter 6 will then discuss the language skills required, thus outlining the construct and content required for the assessments and their corresponding grading criteria. The discussions in Chapter 6 will be divided into the 'standard' English language skills of listening, reading, writing and speaking, although the distinction between these skills will not be wholly clear cut owing to the integrated nature of language for academic purposes.

The EAP Programmes at MUC

Before getting into the discussion, however, a brief recap of the EAP programmes at MUC is required. As explained in the introduction, the Intensive English Programme runs for 13 or 26 weeks from September to January or February to September. Students who achieve IELTS 5.5 or equivalent, with no less than 5.0 in any of the four language skills are offered a place on the 13-week programme; those who achieve IELTS 4.5 in all four language skills are offered a place on the 24-week programme. In the summer, students with an IELTS 5.0 profile, with no less than 4.5 in any skill can study on the 10-week programme. For students who have an IELTS 5.5 profile, with no skill below 5.0 can study the six-week programme. In order to move onto their main programmes of study at the end of these pre-sessional courses, students must pass all assessments with an overall score of 60%, with no skill below 50%. This is deemed equivalent to a high B2 on the CEFR.

In terms of the programme learning outcomes, both the IEP and the SEP list the following learning outcomes:

- To provide English language training skills and practice, and to provide a sound basis for future studies at [MUC].

- To equip students with the necessary academic skills to study up to postgraduate level in a UK higher education establishment.

- To provide experience of being taught English language in a UK higher education establishment.

- To encourage participation in a diverse range of teaching and learning experiences

- To encourage students to use communication skills in a higher education setting and in social situations.

- To enable students to develop knowledge and understanding of approaches to study / culture in a UK higher education establishment.

By way of summary, the IEP and SEP courses aim to help students adapt to academic life at their new institution, whilst also developing the language and academic skills required for coping with HE level study at MUC.

Reporting the focus groups

In terms of reporting the focus group findings, participants are identified by their initials only. As explained in the methodology, in total there were five focus groups, with a total of 20 participants. Table 8 below presents the participants in each focus group:

Table 8 – Focus Group Participants

Focus group	English Development Centre	Subject Specialist – Group 1	Subject Specialist – Group 2	Student – Group 1	Student – Group 2
Participants	LM CL SD BW	DA CW ST IJ RB TM AL	SE AH MW KB EC	OM DY	CY NX

CHAPTER 5

PEDAGOGICAL ACTIVITIES AND TASKS, THE TARGET LANGUAGE USE SITUATIONS AND THE INSTITUTIONAL CULTURE

5.1 Introduction

In order to establish whether the EAP assessments in use at the time of study were useful, it is first necessary to define the specific needs and goals of the departments in which the students will be studying after progression from the EAP programmes (Ypsilandis and Kantaridou, 2007). Additionally, as Lebcir *et al.* (2008) explain, a key factor in success for international students is "adjustment to a new academic environment and culture" (p. 269), since they have to learn to cope with a potentially vastly different approach to teaching and learning. Further, the assessments with which they are faced as part of their programmes of study could be very different to those they have experienced in their home countries (Cheng and Erben, 2012; Bacha, 2010; Lebcir *et al.*, 2008; Kember, 2000). With this in mind, this chapter will use the findings from the assessment audit as well as the focus group discussions to establish three main things: the PATs required of students at MUC; the TLU situations in which they will find themselves; and finally, MUC's academic culture.

I will start by considering classroom expectations, looking at teaching and learning styles and what is expected of students in the classrooms. This will consist of an amalgamation of the findings from the subject specialists and the student focus groups, with some additional comments from the English teaching team. Following this will be discussion of what is expected of students outside the classroom in terms of their learning; the final section of the chapter will look at the results of the assessment audit, along with comments from both staff and students, so as to ascertain the typical assessment tasks which students will face during their time at MUC. Throughout the

discussion, I will comment on how the EAP assessments which were used between 2010 and 2014 did or did not reflect these activities and skills, or whether they offered the students the opportunity to develop them, with the aim of this chapter being to establish to what extent the assessments were authentic. This will involve considering how well the assessments reflected the tasks discussed, how these are operationalised in the grading criteria, and taking into account their face validity in terms of presentation. In addition to considering the face validity, this section will also go some way to establishing the constructs of *academic* listening, reading, writing and speaking.

5.2 The PATs and the TLU situations

As is outlined in much of the literature, there are various academic 'tasks' which students are expected to be able to perform during their time at university, and although these may differ slightly from institution to institution, they are often broadly similar in nature. This first section of this chapter will consider these PATs as they are experienced at MUC, thus outlining the various TLU situations at the institution.

5.2.1 The PATs and the TLU situations in the classroom

The most obvious TLU situation, and the one which has the greatest impact on students since it is how they spend a large part of their time, is the classroom situation. As with all other institutions, students at MUC find themselves in a variety of classroom situations, listening to lectures, participating in discussions or even leading seminars and giving presentations. In terms of what is expected in these classroom situations, the first, and probably most common, activity is listening and making notes in lecture situations. During the focus group categorisation activity, all three stakeholder groups listed listening and making notes as a key pedagogical task. This reflects Benson (1994, p. 181) who describes lecture listening as the "central ritual" of academic life in the

West, and Turner (2004) who adds that university pedagogy in the UK still uses "most prominently the lecture (listening and note-taking)" (p. 102).

In further detailing what is expected of students in a lecture situation, the EAP literature provides various explanations of the skills. Lynch (2011), for example, comments that students are expected to listen to lectures and take notes simultaneously, and Taylor and Geranpayeh (2011) take this a step further, explaining that lecture listening involves "interpreting multi-modal input" (p. 91), with students expected to listen and make notes (possibly on printouts of slides) and look at visual aids being used by the teacher. During the discussions with the different stakeholder groups, all groups commented on the importance for students of being able to note down the key points of a lecture for use at a later date. Just how students go about doing this, however, was a matter of some debate.

Starting with the comments from the English teaching team, the majority of the participants were of the opinion that the task involved listening and making notes on a blank page, as highlighted by Wong and Thang (2008), Lynch (2011) and Taylor and Geranpayeh (2011). SD in particular highlighted the importance of being able to "pull out the main ideas from a lecture" (SD, English Development Centre). In contrast to this, however, BW, who has experience of supporting students with specific learning needs in classes disagreed. Thanks to his experience of classroom situations at MUC, he was able to explain that

> For me, from my own experience having sat through a lot of lectures, I mean, no-one, I mean they don't take notes. They're given PowerPoints and they have to fill in the odd word.
>
> BW (English Development Centre)

This was supported by RB in the first subject specialist group who explained that note-taking in lectures at MUC required little from students in terms of being able to make

notes on key points with no support, stating "we do such copious handouts" (RB, Subject Specialist Group One). This was also alluded to by EC and MW who both mentioned that lecture notes were often placed on the university virtual learning environment (VLE) prior to classes. The students made similar comments relating to the situation in the classroom, with NX and OM mentioning that lecture notes were given out at the start of classes or made available on the VLE before sessions.

In terms of describing the TLU situation, what is emerging here is that the standard pedagogical task in the classroom at MUC is not what most of the English team expected when creating the listening assessments. Although there are sessions in which students are not provided with printouts, for the most part, instead of requiring students to listen and note down the key points in a class, the majority of the lecturers appear to provide students with a scaffolded structure. This scaffold is provided to support the students, but it does mean that rather than just listening and noting down the key points as they hear them, lecture listening at MUC involves complex multi-tasking whereby students have to listen, look at the lecturer or the screen/board, read the notes and write simultaneously. This reflects Song's (2011) findings which explain that more often now students find themselves listening to "a lecture based on teachers' guided notes or handouts (sometimes accompanied by visual aids such as PowerPoint slides)" (p. 69).

To bring in the evaluative framework now, authenticity is seem as a key area of usefulness, and thus it would seem that reflecting the tasks most often encountered in life outside the English classroom would make for the most useful listening assessment, Considering the task authenticity in the listening assessments, the two English pre-sessional programmes assessed listening slightly differently when it came to lecture listening as can be seen in figures 4 and 5 below:

Part Three – Listening Section 3

Listen to this lecture on managing a restaurant. You will hear it once. As you listen, write notes in the space below. After listening to the lecture, you will have time to write a summary of the lecture. You will use these notes and the summary to answer comprehension questions about the lecture.

Figure 4 - SEP Listening Task 3 Instructions

Section One – How People Learn Best
Listen to the lecture and complete the framework. You will hear the lecture *twice*. Marks will be awarded for content (3) and abbreviations and symbols (2). You will use these notes to answer comprehension questions at the end of the listening exam.

Def. of multiple intelligences:

Traditional way to measure intelligence:

Reasons some people don't like this approach:

9 intelligences:

2 effects of multiple intelligences theory in classroom:
1)

2)

Example:

Issue with standardized test:

Figure 5 - IEP Listening Task 1

In terms of authenticity, that is how well the tasks reflect the TLU situation of listening to a lecture and making notes, MUC's culture is not wholly reflected in either of the exams. As explained above, there are situations in which students would be expected to listen and make notes without a scaffolded structure, thus the task in the SEP exam is

143

more reflective of this and would provide students with a more authentic experience. The IEP exam is more problematic in terms of authenticity. Although the students are provided with the scaffolded structure, it would be a more useful assessment of their listening skills in terms of authentic representation if students could be provided with PowerPoint slides, rather than a skeleton outline. In terms of usefulness, then, neither exam could be considered a truly useful assessment of the listening skills required for life at MUC since neither exam wholly represents the reality of lecture listening at the university.

As well as listening in lectures, there are several other classroom situations in which students will often find themselves at MUC, all of which further define the TLU situation of the classroom. These include participating in discussions, leading and participating in debates and seminars and presenting ideas to the class. Clearly, these all involve both speaking and listening skills, and are tasks which could, and should, be reflected in any authentic assessment schedule.

As indicated in some of the literature, classroom situations in Western universities can be very different to those experienced by students in their home countries. Bygrave *et al.* (2014) for example explain that many international students find it difficult to adapt to this new teaching style when they first move to another country to study. The interactive approach favoured by tertiary institutions in Western countries is also considered a particular challenge for students by some authors (for example Smith and Khawaja, 2010; Andrade, 2006), and this point was reflected by student participants DY and OM during their discussion. OM explained that one of the key speaking tasks she believed necessary for coping with studies at MUC was asking questions in class. DY in response to this, explained the cultural adjustment element of this as highlighted by

Evans and Morrison (2011), Kingston and Forland (2008) and Kember (2000). She stated:

> As I said, the speaking, some of us especially from Asian countries, some might, some people may be worried about speaking. They have no confidence in that, but although the teacher told us to speak loud or don't be afraid to make mistakes, we are worried about speaking, especially making mistakes.
>
> DY (Student Group One)

In the second student focus group, CY also recognised issues faced by his classmates in terms of speaking up in class explaining

> The problem is not the, I think the problem is because they don't wanna speak. They fear it. Now it's quite fun but I have to say you'd better change get the brave heart instead of helping them to speak
>
> CY (Student Group Two)

These two comments demonstrate both the recognition of the importance for participating in class, as well as the reluctance to do so. Although this reluctance could be as a result of language issues or a lack of confidence in speaking, what these quotes demonstrate is that international students recognise that there is an expectation to participate, as well as a reluctance to do so. This reluctance to participate, be it due to language issues or cultural adaptation issues, could result in the potential problems which students have with the more interactive teaching style which were reported by both Andrade (2010) and Ramachandran (2011). Ramachandran (2011) emphasises the "limited participation by international students in classroom activities" (p. 208). Andrade (2010, p. 223) also considers this, but from a slightly different angle, explaining that, in her study at least, for international students:

> classroom discourse patterns present challenges as these differ culturally and [students] may be uncomfortable with what they view as [home] students' lack of respect for the professor.

Looking at comments made by the staff in regard to this, CW in the first subject specialist group explained that he expects students to participate in classes, highlighting that he wants students to ask questions when they do not understand. Expanding this

later in the discussion, he recognised the potential cultural problem, reflecting

Andrade's (2010) comment above. He explained:

> …also in some cultures as you know it's not good to talk, it's not good to ask a
> questions, because somehow you're insulting the lecturer by implying that they
> haven't, that you haven't told me everything I need to know.
>
> CW (Subject Specialist Group One)

In the second subject specialist group, KB also commented on speaking, explaining that

an important aspect for her is "confidence in sharing speaking and sharing opinions"

(KB, Subject Specialist Group Two), something which Cheng and Erben (2012) as well

as the student participants in the current study explain is new to many international

students.

With regard to the authenticity element of the framework, in terms of the assessments,

no matter how authentic a speaking assessment task is, it is unlikely to re-create the

situation of being in a classroom and asking questions, thus authenticity is not a viable

option with regard to this particular activity. Acculturation here is more likely to be

achieved through the taught sessions on the English programme rather than the

assessment schedule. In terms of the authenticity element, what can be achieved

through careful assessment design is an opportunity for students to gradually build the

skills required for participating in class, such as putting forward an opinion or agreeing

and disagreeing with the points made by other students.   Ferris and Tagg (1996a)

recommend that students be provided with opportunities to build their skills and

confidence in speaking and sharing opinions and these skills are developed at various

points during the IEP exam schedule. This programme includes three relatively

informal speaking tasks, the digital portfolio, which students produce in their own time,

record and email to their lecturer for grading. The remits for these three assessments are

presented in Figures 6 to 8 below and show how the students are given the opportunity

to slowly build their group participation speaking skills.

Working with a partner, discuss the topic of managing time and possible solutions to the demands of this. Each of you will discuss some of the issues you have encountered.

Prior to discussion, **research the topic** (either alone or together) and **make a list** of **suggestions** you would like to mention related to the following points

- What causes poor time management?
- What experiences have you had of poor time management?
- Possible solutions to some of the issues.
- Advice for students coming to study at university in the UK

You must record your discussion and send it to your lecturer.

Remember to:

- *Use expressions to agree/ disagree/ give opinion*
- *Explain and rephrase where needed*
- *Ask for clarification / more information if you are not sure*
- *Make relevant comments in response to your partner*
- *Engage your partner in the discussion*

Figure 6 - IEP Semester 2 Digital Portfolio 1

Working with a partner, choose one of the following **global issues:**

- The use of fossil fuels
- Overcrowding
- Waste
- Increasing use of cars

Before you speak, each make a list of the **consequences of** and the **solutions to** this problem.

Discuss the consequences of the problem and put forward possible solutions. You should also **evaluate** these solutions (e.g. the advantages and disadvantages – think about different **perspectives**; what needs to be done to make each solution work; whether a solution could create further problems). **If possible,** try to agree on a solution.

You must record your discussion and send it to your lecturer.

Remember to:

- *Use expressions to agree/ disagree/ give opinion*
- *Make relevant comments in response to your partner*
- *Engage your partner in the discussion*

Figure 7 - IEP Semester 2 Digital Portfolio 2

Working with your group, discuss the following questions about **cause and effect relationships**:

- Does poor health lead to poor economic circumstances?
- Do poorer economic circumstances lead to poor health?
- Does a better education lead to a better economic situation?
- Does a better economic situation lead to better education?

Before you speak, each make a note of your **responses and reasons** to each questions.

Discuss the relationships. **If possible,** try to agree on the cause and effect relationship in each case.

You must record your discussion and send it to your lecturer.

Remember to:

- *Use expressions to agree/ disagree/ give opinion*
- *Use language to reach a conclusion*
- *Ask for more information if needed*
- *Use cautious language*
- *Make relevant comments in response to your group members*
- *Engage your group members in the discussion*

Figure 8 - IEP Semester 2 Digital Portfolio 3

All of the tasks presented above are graded using the same generic grading criteria. In terms of how these operationalise the skills related to speaking in classroom situations, students are graded as seen below in Table 9:

Table 9 - Speaking Grading Criteria Operationalization of Classroom Speaking Tasks

Percentage grade	Descriptor
85 – 100	Uses range of expressions (giving opinion; interrupting etc.) with complete confidence. Excellent level of contribution displayed.
75 – 84	Uses range of expressions (giving opinion; interrupting etc.) with confidence. Very good level of contribution displayed.
60 – 74	Uses range of expressions (giving opinion; interrupting etc.) with some confidence, although these might be limited in scope Good level of contribution displayed.
50 – 59 (borderline fail)	Uses expressions to give opinion; agree/disagree; interrupt although these may be limited in scope and repetitive. Adequate level of contribution displayed.
40 – 49 (fail)	Limited use of expressions to give opinion; agree/disagree; interrupt Average level of contribution
35 – 39 (fail)	Rarely uses expressions to give opinion; agree/disagree; interrupt Low level of contribution displayed
Below 35 (fail)	No attempt made to contribute/interact with other students

Although only limited in their description of classroom participation speaking tasks, these criteria do at least partially support students in their development of the speaking skills required to cope with in-class speaking activities, particularly the expressions of interrupting, giving opinion and agreeing and disagreeing. These criteria, however, could be deemed limited, and in turn the assessments could lack in terms of their authenticity since they only cover language related to agreeing, disagreeing and interrupting, they do not comment on students' abilities to make suggestions or recommendations or cause and effect. Hence, although the tasks provide students with opportunities to develop 'authentic' speaking skills, the reliability of the grading criteria in how they operationalise the language and skills required by the assessment tasks is questionable.

A further point of note regarding speaking tasks is that the digital portfolio approach is not reflected in the SEP assessment schedule, which implies a gap in the assessments that I believe undermines the usefulness of the speaking assessments used on this course. Since these are supposed to be equivalent programmes, with the same programme learning outcomes, differing mainly in when they take place, the assessment schedule should also be equivalent, where possible.

5.2.2 PATs outside the classroom – the wider academic skills

So far, I have considered the in-class TLU situations, but as is widely known, students are required to carry out additional learning activities outside the classroom, in the form of research, reading and group work. In order to be able to carry out these tasks, students need not only to be proficient in English, but also aware of the demands and requirements of these particular tasks. It is these tasks which some of the literature suggests are likely to require the largest adjustment for students from countries such as

China, where teachers are more likely to provide students with all the information they need to be able to pass their exams (Kingston and Forland, 2008). Ramachandran (2011), in fact, highlights that "many [international] students depend on their professors for study material and do not see reasons to develop their own materials from a variety of authentic sources" (p. 208). Although the research presented in Chapter 3 demonstrates that not all international students will find this a difficult or even alien concept (Ha, 2006; Kumaravadivelu, 2003), the research from the focus groups shows that this need to research into a topic is expected of students at MUC, and it is something which the participants believed was problematic for international students.

Before looking at the assessments in more detail, an area which was mentioned by several participants across all focus groups was the idea of autonomous learning, reflecting the comments from Kingston and Forland (2008) and Ramachandran (2011) above. In the English team discussion, the sticky note activity produced a study skills category and in the discussion which followed, SD emphasised working independently and autonomy as key skills for coping with Western academia, explaining, however, that "it has to be directed, at least initially [because] autonomy does not come, it will not come easily" (SD, English Development Centre Focus Group). This idea of learning independently, or being autonomous, was also picked up by the second subject specialist group with both AH and EC commenting on the importance of students being able to work independently.

In applying the framework and considering task authenticity, in terms of the assessments, autonomy is not something that can necessarily be directly assessed, however it is something which students can develop whilst working on an effectively designed assessment. Considering the English assessments on the IEP, the speaking

assessments, as detailed above, require students to record discussions in their own time to be sent to lecturers. This entails a certain level of autonomy and time-management on the part of the students in terms of deciding what to discuss, when and where, how to record it and how best to send it to their lecturers. Although only minimal in terms of independent learning and time management, these small steps still provide students with opportunities to begin to work at their own pace, and make their own decisions about the contents of their work. As implied by Kingston and Forland (2008) and Ramachandran (2011) above, such skills may be new to many, although not all, international students.

As well as the speaking assessments, the writing assessments on the IEP require students to work semi-autonomously, although they are guided through the process by their lecturers. Figures 9 and 10 below contain the instructions for the IEP writing tasks:

Module Title
Writing

Assignment Title
Students and Part time work

Hand Out Date:
Week 1

Hand In Date:
Section A Week 5
Section B Week 12

Assignment Remit
More and more students are now working part time during their studies. Using the articles in your reading pack, discuss the good and bad points of working part time whilst studying.

Section A (20%)
For this section of the assignment, you need to create a plan for your essay. This needs to include a time-line, explaining how you will organise your writing and research. This needs to be a realistic plan which you will try to follow during the semester. When you are writing your plan, look at the Module Delivery Scheme for information on important hand-in dates. You also need to produce an introduction to your essay of approximately 100 words, and a supporting paragraph plan which outlines what you will discuss in each of the paragraphs of your essay.

Section B (80%)
For this section, you need to write your completed essay. You will need to answer the question set in the assignment remit and write 1000 words in essay format. Use references and quotes from the articles in the reading list to support your discussion. You cannot use any other articles for this assignment. You will need to include a new introduction with corrections as suggested in your feedback for Section A.

Figure 9 - IEP Semester 1 Writing Assignment

Figure 10 - IEP Semester 2 Writing Assignment

In considering the authenticity area of the framework, the independent learning and time-management skills are developed in several ways through these assignments. Firstly, there is the amount of time which students are given to carry out the work; they have to plan this time carefully so as to be able to collect the data they need to produce the work; this reflects what they will encounter when they move onto their main programmes and so in terms of authenticity, these assignments could be considered useful. Despite the authentic nature of the tasks in terms of time management, this is, however, guided quite carefully in the semester 1 assignment with the help of the time plan and the essay plan. Additionally, in the second semester assignment, students are required to design and carry out a small-scale primary research project, making their own decisions about how best to do this and when to do it. As is recommended by SD in the comments above, these first steps into autonomy are guided by the lecturers throughout the semester, although it is still ultimately the decision of each individual student how to approach the task. In terms of authenticity with regard to autonomous

learning and time management, then, these assignments are useful. Where the assignments may be seen to be less authentic is with the reading lists. In an attempt to reduce plagiarism, and bearing in mind the fact that these assignments are essentially assessing students' writing skills, the students are restricted in their reading choices. Thus, in this area the assignments are less authentic, and do reduce students' chances to further develop as autonomous learners.

The SEP writing assessments also provide students with the opportunity to work on their planning and time-management skills (see Figure 11 below).

Module Title
Writing

Assignment Title
Analysis of a hotel's online resources

Hand Out Date
Week 1

Hand In Date
10-Week students: Week 9
6-Week students: Week 6

Assignment Remit
Online resources (for example, websites) present hotels with new opportunities to market themselves to customers. Use the AIDA model to evaluate the online resources of **one** of the following hotel chains:
  (1) Accor Hotels
  (2) Premier Inn
  (3) Hilton Hotels
  (4) Novotel Hotels

You need to evaluate the resources in relation to the three following scenarios:
  a) A family looking for a weekend break
  b) A business looking to hold a conference
  c) A couple looking to visit the restaurant for a meal

You can select appropriate diagrams and pictures to use in your literature review and analysis section if appropriate. Make sure that you provide clear references for any pictures and diagrams, though.

IMPORTANT!
Please note – you can **only** reference materials included in the reading pack. You do **not** have to read every article from beginning to end. Instead, you should read **relevant** sections of the articles in order to help you write your report. You will need to identify what is relevant and what isn't.

Format
1,500 word individual report based on input from classes and individual research.

Figure 11 - SEP Writing Assignment

As with the IEP assignments, this task provides students with the opportunity to produce work over time, thus giving them practice in planning and time-management, with even the six-week programme students having to produce work over a period of time which is similar to what would be expected of them when they progress (typically, assignments are done over four to twelve weeks).

Having considered the need for time-management, planning and autonomous learning, all of which are key for coping with UK HE, the next area for discussion is researching for assignments, an area which might some literature implies is a new skill for many international students, particularly those from East Asian cultures. Both Ramachandran (2011) and Jordan (2002) comment on the fact that many international students expect their lecturers to provide all reading material, with Ramachandran (2011) highlighting that students "do not see reasons to develop their own materials from a variety of authentic sources" (p. 208). As is evidenced further in the literature, and also in the focus groups, however, it is quite clear that at MUC, as with many other institutions, students are expected to carry out research so they can complete assignments. As Jordan (2002) explains, it is expected that students will "assume responsibility for finding self-selected readings from library resources relevant to their academic interests and needs" (p. 108). Additionally, the lecturers expect students to approach the sources critically from the outset (Smith and Khawaja, 2010). Although, as is indicated in the discussions in Chapter 2, the deficit model is not necessarily the most helpful approach to take when discussing international students, as the discussions in this section progress, it will become clear that it is a belief which permeates the culture at MUC.

Coming to the primary data, for the English teaching team, reading and researching was mentioned on several occasions, and it was discussed by all of the participants. In fact,

this group dedicated an entire category to finding information, which they titled

'Digging Out'. SD summarised this category, explaining:

> I think for me research skills involves being able to use databases and the library and being able to pick out resources … For me research is about getting the information … It's digging it out, isn't it? So, can you get into the database?
>
> SD (English Development Centre)

He later added "I've got using the VLE … coz there might be articles on it… because

lecturers put articles on there" (SD, English Development Centre). While discussing

the same category, LM stated that for her, what was important in terms of researching

was "selecting key words for research, that's like actually deciding what your key

words will be to do your research" (LM, English Development Centre). As well as

discussing the idea of key word searches, a critical skill in itself, LM explained

> Read critically, you see I've got effective reading skills and that's what I meant by that, erm, critical thinking, maybe.
>
> LM (English Development Centre)

The area of critical skills was, in fact, afforded its own category by the English team,

and BW summarised this category by explaining:

> Critical stuff here it's about the much talked about, basically…, reading critically and yeah, awareness that there are other arguments out there.
>
> BW (English Development Centre)

CL expanded this, adding:

> But I think it's also critical stuff, recognising the different types of, er, things that you're reading and things that you're thinking about, well, what are we going to do with it? We've got something that's non-academic, but how can we evaluate it? How can we be aware of how to use it?
>
> CL (English Development Centre)

So far, then, the English team are of the collective opinion that students need to learn how to search for sources (the technical skills of finding information in a library, online database or on the university VLE), and then approach these critically to ensure that they are suitable for their particular research purpose, and the following comments from students and subject specialists demonstrate that these two groups share this opinion.

Beginning with the students, from the outset of the first student focus group, OM suggested a research skills category. She explained "how to use [the VLE]... was so difficult for me" adding, "researching step by step" (OM, Student Group One) was something which she has needed to complete her studies. DY, in the same discussion, explained:

> In my course now I'm studying the research. Academic research is different from my bachelor or high school … especially the research skills, it's new.

> DY (Student Group One)

In the second student focus group, both participants also commented on research. CY started the discussion stating:

> Information searching … Reading the news for … seeking some relevant information… And explore a case related to the lecture.
> CY (Student Group Two)

In response to this, NX explained that finding information for assignments was a key skill.

What these comments demonstrate is that for the student participants, the idea of having to search for their own materials to help them to prepare an assignment was new, and something which they have had to learn to do.

Coming to the subject specialist staff discussions, the need to research, and to respond critically to texts, was highlighted by all 12 participants. In the first focus group, all seven participants commented on both the critical element and the research skills and during the sticky note activity, this first group had a "critical thinking" category, as well as a researching category. Notable comments from this group included those from AL who explained "I've got differentiating theory to anecdotal theory … and analysing" (AL, Subject Specialist Group One). With particular relation to critical skills, ST recognised the potential difficulties facing many international students, explaining:

> Where I think our expectation is that they should be able to come in and critically analyse I think experience shows that that's not necessarily the case and we then have to give them to tools to be able to reach that.
>
> ST (Subject Specialist Group One)

In the second staff focus group, there was the following exchange at the start of the sticky note activity:

> MW: I've got planning … how to do a proper lit search in the library, research skills, analysis and evaluation, reading critically and being able to draw proper conclusions from your work.
> KB: Being able to speed read, coz that, they struggle with that, and then being able to take notes quickly as well. To think outside the box, be creative, and to be able to question what they read.
> EC: Now I've put accessing reading. Could we put research skills on there, it comes in, doesn't it?
> AH: Research skills. I've called that differently: Research and finding information.
> EC: because a lot is online I put IT skills because without IT skills you can't access anything, so I don't know with that. I don't know whether this should come in is working independently.
> SE: Ok. Sourcing references. Is that research?
> AH: Referencing and supporting ideas.
> SE: I've got here understanding research methods, which I guess is research skills as well.
>
> Extract from Subject Specialist Group Two

There is mention here made of the IT skills, and the need to be able to physically carry out the search for literature, be it online or in the library, which reflects the comments from SD in the English team discussion. In addition, however, what is clear from the

above exchange is the importance placed on research and critical thinking by the subject

specialist staff, something which was picked up again later in the discussion as is seen

in this exchange:

> KB: How about being creative? I've said to think outside the box which is, coz a lot
> of international students, particularly Chinese, I'm being stereotypical here, but
> they're very restricted in how they express views.
> EC: ...they're not kind of very...
> SE: Innovative!
>
> Extract from Subject Specialist Group Two

The need for students to be creative, for them to "think outside the box" was clearly

important for this group, and seen as lacking, reflecting the deficit model as commented

on by Andrade (2006) who explains that tutors reported that students "lacked critical

thinking skills" (p. 138) and were often "unaccustomed to analysing the strengths and

weaknesses of an argument" (p. 139). SE, in fact, took this idea a step further,

commenting that:

> If they look at research as a practice or an academic skill, erm, research I think can
> quite often be interpreted as a researcher sitting in an office looking at, you know,
> Encore or whatever research app, whereas research actually in context is
> something quite different. ...research takes a lot of forms, actually.
>
> SE (Subject Specialist Group Two)

This idea of working in context, in particular case study work, reflects the comments

from the student participants, three of whom mentioned case studies as a key form of

assessment used at MUC. Researching in context is something which carries through

all levels of study at MUC, with the majority of assignments providing students with a

specific context around which to focus their work. Thus, a vital part of the critical

approach required of students at MUC is ensuring that the research carried out remains

focussed on a particular case, or within a particular context. When I consider the

English assessments below, I will return to this idea of research in context since it is an

important element of academic life at MUC.

What emerged from all of the discussions was that the practical skills of research – that is actually finding useful information – and the more abstract critical thought – deciding which search terms to use, where to look, which sources are the most suitable and researching with a particular context in mind - are considered key, which supports the findings from the literature. The research consulted for this study, however, does not comment on the idea of research in context. Thus, in terms of evaluating the assessments, key PATs at MUC are researching in context, finding sources and making critical decisions about these sources. As ST states above, students need to be given "the tools", so the assessments need to assess, or provide students with the opportunity to develop, the research and critical skills, with a particular institutional focus on using assignments which provide a particular context. Coming again to the authenticity element of the evaluative framework I will now consider the writing, speaking and reading assessments since there is an element of both research and critical thinking in each of these assessments.

Starting with the speaking assessments, all of the speaking tasks provide a context for the students to work in. Take the following examples two seminar tasks and a digital portfolio task (Figures 12 to 14 on the following two pages), which all offer students guidance on the context.

Figure 12 - SEP Speaking Assignment

Figure 13 - IEP Speaking Assignment

Working with your partner, you are going to discuss the problems of homesickness, and possible solutions to it. Each of you will discuss some of the issues you have encountered as a student.

Before you speak, each make a list of suggestions you would like to mention related to the following points:

- What causes homesickness
- What experiences you have had of homesickness
- Possible solutions to some of the issues
- Advice for students coming to study in the UK from your country

During the discussion (6-8 mins) try to help each other

You must record your discussion and one of you must send it to your lecturer.

Figure 14 - IEP Semester 1 Digital Portfolio 1

These are helping prepare students for the idea of working within a specific context, asking them to offer suggestions or recommendations about a particular issue based on their reading and their own experience. Each of the assessments, although slightly different in nature, provides a clear topic for the students to research and asks for students to research this topic with a particular purpose in mind, be it the problems of Ibis Hotels' online resources, the pros and cons of university students working part time or the issue of homesickness for students moving to the UK. These clear contexts provide students with guidance similar to that found on other assignments given to students at MUC thus offering them an authentic assessment experience and so, in terms of the framework, demonstrating the usefulness of the tasks. A similar approach is taken in the writing assignments, as can be seen in Figures 12 to 14 above. Additionally, the context requires students to approach their research from a particular point of view, encouraging a critical approach to the research information which needs to be evaluated to ensure that it is suitable for that particular context.

In addition to providing a context, thus providing students with a developmental opportunity in terms of critical reading, the writing assignments also directly assess students' ability to think about sources critically as can be seen in Figures 15 and 16 on

the following page. The literature matrices required by the assignments perform three functions. Firstly, they provide students with the opportunity to practise producing something which is a mandatory aspect of dissertation writing at MUC: the literature matrix. Secondly, they offer the student another opportunity to practise critically evaluating the literature; and thirdly, they give the English teachers direct evidence of the students' ability to do this. The authenticity of these tasks further enhances the usefulness of these assignments in helping students to prepare for their future studies at MUC.

The final assessments which include a critical thought element are the reading assessments. The SEP exam, as is evidenced in the extract below, tests critical thinking in task five. This particular task tests critical thinking in that students are required to identify the key words in the task they are 'preparing' for, and then, using the information in the text, create a series of questions which might be answered when reading the text. In order to be awarded the marks, however, the questions must be so specifically written that they reflect both the question and the text they have been given. According to the mark scheme, students are not to be awarded marks for writing questions which relate to *just* the essay question or *just* the text; their questions must include elements of both. Therefore, a question such as 'what are the main features of hotel websites in Europe and Asia?' would not be suitable since it relates only to the essay question. Similarly, it is not acceptable to ask 'what managerial suggestions are proposed?' since this relates only to the text. This reflects the requirements of the subject specialists that students take a questioning approach and are able to focus their research on a particular context.

Part 3: Reading passage 3

Imagine that you are writing a report on **the differences between hotel websites in Europe and Asia.** You have found this article about Taiwanese and Chinese websites, published in 2012 in the Journal of Travel & Tourism Marketing.

> **What Does Hotel Website Content Say About a Property – An Evaluation of Upscale Hotels in Taiwan and China**
>
> ABSTRACT
>
> Websites are important platforms for hotel marketing. Website content can directly influence customer preferences and decisions. This study proposes a comprehensive model, which is more objective and valid than others, to evaluate 158 hotel websites in Taiwan and China according to three dimensions: feature breadth, stage of website development, and feature enrichment. The samples came from four different types of hotels: Taiwanese international tourist hotels, Taiwanese ordinary tourist hotels, Chinese five-star hotels, and Chinese four-star hotels. We find that these four hotel types emphasize features on their websites differently and that the degree of website robustness ranks as follows: Taiwanese international hotels, Taiwanese ordinary tourist hotels, Chinese five-star hotels and Chinese four-star hotels. Managerial implications and suggestions are also proposed.

Task 5

Using the title and the abstract, think of **2** research questions to help you get useful information from this article for your report.

_____
_____
_____
_____

Figure 15 - SEP Reading Exam Task 5

In the IEP reading exam, students are provided from the outset with opportunities to demonstrate their critical reading skills. Figure 16 below includes the instructions for the first two tasks.

Task A
Read the essay question in **bold** below and <u>underline</u> the key words.

What are the positive and negative effects of globalisation? Discuss

Task B
Read the three abstracts below and answer these questions:

1. Which journal article is unlikely to help you to answer the essay question in task A?
_____
2. Why is this article not likely to help?

Figure 16 - IEP Reading Exam Tasks A and B

These first two tasks clearly test critical reading in that students are required not only to read and understand the texts, but also to identify which of the sources would not be useful, and explain their reasons why this is the case. Again, students are asked to focus on a particular context from the outset. To consider the evaluative framework, in terms of usefully testing critical reading skills, then, the tasks here reflect the PATs as outlined by the focus group participants.

As for the technical search skills, that is using the library and online searches to find sources of information, there is very little opportunity for students to develop these skills since in order to ensure they are accessing materials which are of sufficient academic quality, and to account for the short time during which students can research whilst also working on improving their language level, they are provided with a reading pack. This also ensures minimal plagiarism opportunities since the staff are very familiar with the reading materials, and students have less opportunity to purchase their work from essay-writing services or copy from online sources. As discussed above, this does, however, impact on the usefulness of the assessments in terms of students learning to find reliable sources.

In terms of the technical search skills, then, in order for the writing and speaking assessments to be more useful to students, they could be revised to be more open, perhaps offering students a shorter reading pack which they then have to supplement with further research. Alternatively, rather than providing the reading in a pre-prepared reading pack, students could, instead, be given a reading list and then shown how to access their own sources. This would enhance the authenticity of the assessment experience for the students, whilst also improving their usefulness since on their main

programmes of study, students are provided with reading lists for all modules and required to search for the sources themselves.

5.2.3 Assessment-related PATs

The most obvious of the assessment-related PATs are the assessments themselves – the essays, reports, dissertations, seminars and presentations - although there are other assessment-related PATs which I will come to later. To consider these, I start with the assessment audit I carried out for the academic year 2013-2014. Although only basic statistics, the figures here provide an adequate snapshot of assessment tasks at MUC. Figure 17 below shows the breakdown of assessment types across the whole university. The labels used for this figure come directly from institutional labels for different assessment types.

Overall Assessment Type Breakdown

- Poster 1.2%
- Role Play 1%
- Dissertation 0.4%
- Portfolio 3.5%
- Case Study 5.5%
- Practical 9.8%
- Report 29.9%
- Debate 2.4%
- Seminar…
- Presentation 17.5%
- Exam 10.8%
- Essay 18.1%
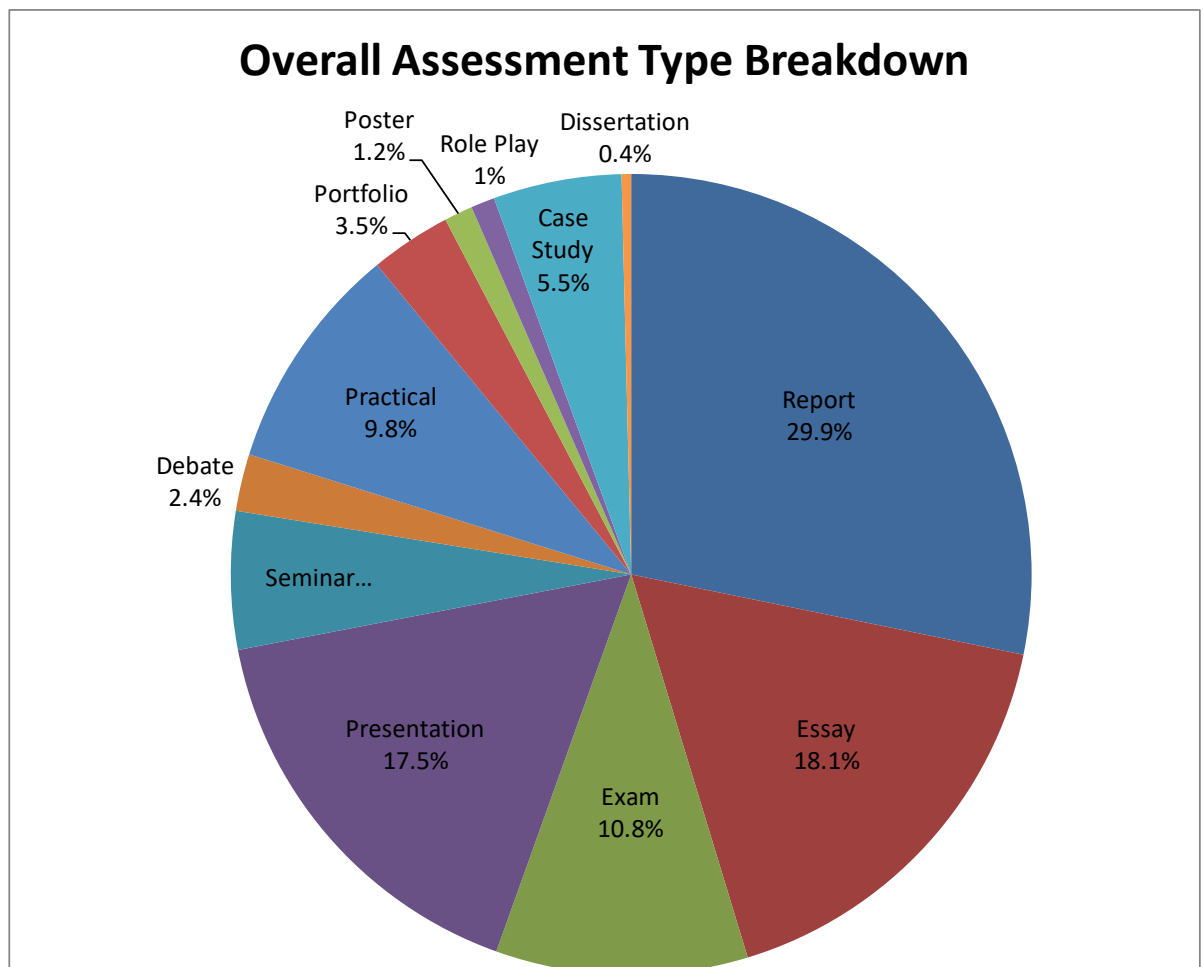
Figure 17 - Overall Assessment Type Breakdown 2013-2014

As can be seen from Figure 17, written work in the form of essays or reports makes up 48% of the assessments used across the university, with written exams taking up just 10.8%. These figures, however, are for the university as a whole, however the majority of students on pre-sessional programmes progress onto programmes at levels 6 or 7, that is the final year of a degree or a postgraduate course, so looking at the breakdown for these levels would be more useful for the purposes of this study. At level 6, the written coursework elements, that is essays, reports and the dissertation, make up a total of 60.6% although at level 7, this falls to 51.4%. At these levels, exams make up 10.6% and 11.1% respectively.

When comparing this to some of the data from the literature, this is a large shift in assessment approach for students coming from countries such as China and Hong Kong (Zhang and Mi, 2010; Kingston and Forland, 2008). Since the majority of MUC pre-sessional students are from these countries, it is safe to assume that this is also the case for our students. Thus, a key part of the academic acculturation process, and consequently a way to make the assessments more useful for these students, is making the students more aware of the demands of these particular assessment PATs.

The figures presented so far, then, highlight the importance of essays and reports at MUC, especially for the levels of study into which the students progress at the end of their pre-sessional programmes. These figures were broadly reflected in the focus group discussions, with both student groups including an academic writing category in the sticky note activity, demonstrating the importance which students place on this activity. The area of academic writing was also returned to several times during both

student discussions, with all four students emphasising its importance. Additionally, they talked about how the approaches in the UK differ from their home country, which reflects the comments from Bacha (2010).

Moving to the comments from the subject specialists, they were also very concerned with students' academic writing, and all twelve participants considered essay- and report-writing to be crucial for coping with HE level study. Their comments, however, did not wholly reflect the results of the assessment audit, particularly in the second group. KB, for example, commented that "in the tourism school, erm, and essays just don't really happen anymore, it's all about reports in context" (KB, Subject Specialist Group Two). Certainly, breaking down the data into more detail, reports are more common than essays across the university, but the essay remains an important academic task at MUC. Taking level 6 as an example, essays make up 26.9% of the assessments used across all programmes, and reports are 32.7%; in contrast for the Tourism department, essays make up 17.9% of the assessments used, whereas reports are 35.7%. Despite these minor discrepancies in assessment type, what is clear from the above data is that both essays and reports constitute key assessment PATs with which students need to be familiar, and which they need to be confident in producing if they are to cope with their studies. Additionally, since the English pre-sessional programmes are for students entering *any* school in the institution, it is still useful to include an essay in the overall assessment schedule if possible.

In terms of the assessments used on the pre-sessional programmes, a combination of essays and reports is used, with emphasis on report-writing for the shorter programmes. The 26-week IEP starts with an essay in the first semester, and then students progress to

report-writing in their second semester. In contrast, during the SEP, students are assessed using a report, since this is the more commonly-used assessment type on level 6 and level 7 programmes (32.7% and 31.9% respectively).

Considering the evaluative framework, to further enhance the usefulness of these tasks, the English team decided to reflect the assessment approaches used at MUC, including mirroring the layout of the tasks. This enhances the face validity of the tasks, which in turn aids the students' institutional acculturation. Despite the positive impact that an authentic task layout might have, it was commented on negatively by some of the subject specialist staff. In the first group, ST and DA discussed the point in great detail, with DA commenting "there's an awful lot of information on these remits" (DA, Subject Specialist Group One). Additionally, in the second group, both KB and EC commented on the fact that the way the assignment remits were structured was "scary", with EC explaining further that:

> For me I think if I was to go for example abroad and to have something similar, if I had a, much like what [AH] is saying, if I had a piece of paper given to me and I wasn't able to understand English, I'd struggle with the paper initially and I'd have a fear factor and that fear factor would put a block on me being able, it would kind of that brick wall would go up.
>
> <div align="right">EC (Subject Specialist Group Two)</div>

In terms of the acculturation, however, ST in the first subject specialist group did discuss just how important it is for students to be able to cope with complex assignment remits, since these are the institutional norm. He explained:

> Well we talked recently you know about students coming from all the way round the world and all of a sudden in week one after induction where we don't usually talk about how the remits are gonna look and things like that, they're bombarded with three or four bits of work and they must think 'Wow! Welcome to [the Midlands]! Fantastic. Welcome to [MUC]!'. And so we are assuming that they know what the remit is gonna look like, what's involved, they look at these learning outcomes they think 'Oh! What on earth are those?' And almost they actually look at the learning outcomes more than the remit themselves coz they think that's where the task starts. They're numbered, the tasks are numbered therefore I've got to deal with those. ... Sometimes you've gotta really find the

assessment in amongst all of the other stuff that you've got. it must be mind-boggling for them.

<div align="right">ST (Subject Specialist Group One)</div>

In fact, in the second group, understanding assignment remits was noted down during the activity stage by two of the participants. Considering both of these viewpoints, then, although by following institutional norms, the English assignment remits might initially appear "scary" to students, by offering authentically presented tasks, with high face validity, the English team are providing students with an important academic acculturation opportunity. In doing this, the impact of the issues highlighted by ST could well be reduced since the students are guided through this style of remit during their English programme. Therefore, if we take one of the main aims of a pre-sessional programme to be academic acculturation, and we recognise assessment as a key PAT at MUC, then helping students to become familiar with the layout of assessments by mirroring this in EAP assessments helps ensure usefulness in terms of the evaluative framework.

It is necessary, however, to consider more than just the written assessment tasks when looking at assessment PATs since written assignments are only one part of the story in UK HE as is widely reported in the literature (see, for example, Ramachandran, 2011; Zhang and Mi, 2010; Campbell and Norton, 2007), and the same is so at MUC. Returning to the assessment audit and the overall assessment breakdown presented in Figure 15 above, spoken assessments – that is presentations, debates, seminars and role plays – constitute 26.8% of all assessment activities at MUC, with presentations making up the largest proportion of these (17.5% overall). Considering levels 6 and 7, spoken assessments constitute 39.5% and 36% respectively, with presentations being 26% and 19.4% of the total assessments used. What is clear from these figures is that

assessments PATs at MUC include a large number of spoken tasks, and in particular presentations (both group and individual).

The importance of presentations was reflected in the subject specialist discussions in particular, and in the first group, five of the seven participants specifically mentioned presentations. Notably, CW explained "a lot of our students have to do presentations" (CW, Subject Specialist Group One), and other participants noted presentation skills down during the sticky note activity. In the second group, MW asked "Do they have to do any presentations? We do a lot of presentations in our school" (MW, Subject Specialist Group Two), thus emphasising the importance of presentations over seminars and debates. This very much reflects the situations reported in the literature, with a number of authors commenting on the importance of presentations in UK HE (see for example Evans and Morrison, 2011; Wong and Thang, 2008; Turner, 2004; and Ferris and Tagg, 1996a).

In looking at the EAP assessments, however, speaking skills were assessed purely through the use of seminar and group discussions during the time of study. The main reason behind this was that it is more difficult for students to read from a script during a seminar than it is during a presentation. Therefore, in order to attempt to achieve a more accurate assessment of the students' speaking ability, seminars were chosen. Further, it is more likely that students will have encountered presentations in their previous education, however seminars and debates are less well-known approaches to assessment (Andrade, 2006), and are therefore using them on the IEP and SEP goes some way to helping students with their academic acculturation.

Looking purely at the statistics, however, in terms of their usefulness to students for acculturation, it could be said that the use of seminars is a less authentic approach since students are more likely to face presentations on their main programmes of study. Despite this, in terms of their usefulness of providing English staff with an accurate representation of how well students are able to speak, listen to and respond to the points of others, the assessments can be considered useful. With regards to increasing the authenticity of assessment experience, and further enhance assessment usefulness, then, what I recommend is the inclusion of a presentation in addition to the seminar tasks. In doing this, the students would be given practice in the two main spoken tasks used at MUC.

The final area for consideration when it comes to assessment TLU situations is that of groupwork. A large number of the modules at MUC include some element of groupwork, be it group presentations, seminars or group reports and essays. Overall, across all modules at the university, groupwork assessments make up 13.2%. However, when considering levels six and seven, they constitute 24% and 22.2% respectively. To look at this another way, at least 40 credits out of the 120 credits which students can get for their final year will come from groupwork projects, that is at least two modules. When on their masters programme, at least one of the modules will include a groupwork element.

To look now at the focus groups, although not a consideration for the English team, and only mentioned by one student participant, the subject specialists mentioned working with groups on several occasions, and it was considered an area of particular importance in the second subject specialist group with three of the five participants mentioning it. In terms of the literature, groupwork is discussed as a new skill by Ferris and Tagg

(1996b), who explain, in relation to giving presentations, that there is a "growing tendency for them to work in pairs or groups rather than alone" (p. 49). Cheng and Erben (2012) also comment on the fact that it is quite a culture shift for many students when they come to study in the UK and find that their grades, rather than being based on examination scores, are often based on their participation in group projects.

From these comments, then it seems that the idea of group grades and working with others is considered quite alien by many international students and its inclusion in an EAP assessment schedule could greatly enhance the usefulness of tasks in terms of authenticity and representation of the likely TLU situations. Despite this, the assessments used on MUC pre-sessional English programmes until 2014 did not use any group tasks; this seems to be an oversight taking into consideration their importance on the main programmes of study at MUC.

5.3 Summary

This chapter has gone some way to extending and applying the analytical framework from chapter four of this study, and as such has partially achieved the sixth and seventh objective. In establishing the institutional academic culture, as I have done here, I have been able to establish the extent to which the assessments reflected this, i.e. their authenticity, and thus start to establish their usefulness.

What has become clear from the discussions here is that all participants, and in particular the subject specialists very much took a deficit approach to the discussions about the students' abilities, with the expectation that the English team would fill in gaps in student learning and knowledge. Certainly, there is an expectation that in completing the pre-sessional language programme assessments, the students will be able

to cope adequately with all of the challenges facing them during their studies. This brings me to my first main conclusion, namely that in order to be useful, the assessments should provide the students with the opportunity to develop a broad set of academic competencies. These academic competencies could be included in a way similar to the graduate attributes which are developed through HE assessments.

In terms of providing an evaluation of the assessments in use at the time of study, the discussions above have demonstrated that they were partially useful for all stakeholders in terms of institutional acculturation, however, they were also lacking in several areas. Firstly, in general the assessments lacked standardisation across programmes. Additionally, with group work playing such a key role in academic life at MUC, the lack of group work tasks in the EAP assessments impacted on their general usefulness. Looking at each skill individually, the listening exams required greater authenticity in terms of how they assessed note-taking skills. The speaking assessments, although providing a good opportunity for students to gradually develop their discussion skills, and providing an opportunity to demonstrate unscripted speaking ability, lack authenticity in that they do not offer students the chance to develop a key speaking skill: giving presentations. The reading and writing assessments provide students the opportunity to develop and demonstrate their competence in critical skills, thus enhancing their usefulness. Further, the writing assessments provide students with the opportunities to develop as autonomous learners, which are adequately scaffolded to aid their development. However, the imposition of reading lists or use of a reading pack for the writing and speaking assignments do limit the opportunities provided for students to develop as successful, critical researchers.

In order for the assessments to be more useful, as the application of the 'authenticity' element of the framework has demonstrated, MUC's English team needs to further develop the assessments in use at the time of study to ensure that they will provide authentic representations of the tasks students will face on progression, thus enhancing their usefulness for the students. If the English team can be confident that students are demonstrating competence in a range of skills that reflect those which they will face upon progression, they can be confident that the predictions they make about how well students will cope with these tasks are valid, and their feedback and recommendations on the basis of grades are reliable and useful for all. Finally, if the assessments aid students with their academic institutional acculturation through their authenticity, then the subject specialist staff who teach them on their main programmes of study can be confident that they only need focus on subject content and can spend less time helping the students to settle in the institution.

CHAPTER 6

PREDICITIVE VALIDITY AND RELIABILITY

6.1 Introduction

Having now established the usefulness of the assessments in terms of the academic acculturation of the international students on pre-sessional English programmes at MUC, the next steps on the evaluative framework require me to consider the predictive validity and reliability of the assessments and their corresponding grading criteria. As is explained in much of the language testing literature (e.g. Bachman and Palmer, 1996; Davies, 2001; Aryadoust, 2013), and discussed in the literature review, when designing a test an important starting point is to define the construct that is to be tested; the same, therefore, must hold when evaluating the usefulness of in-use assessments. Thus, in order to work towards the predictive validity argument being presented here, this second findings chapter will use the focus group findings in conjunction with information from key literature relating to EAP to establish the different constructs being tested, namely listening, reading, writing and speaking for academic purposes. The validity argument, however, is not just determined through the construct, and so these findings will also be used to establish the specific academic context in which the students will be studying, focussing on how the authenticity of task, or experience, can help predict how well the students will cope with this. The discussions will also consider the contents of the assessments, so as to establish content validity. Additionally, the reliability of the mark schemes and grading criteria will also be investigated.

Since, as Watson Todd (2003) explains, "the main goal of EAP is for students to communicate in academic environments" (p. 149), these skills form the basis of the needs analysis which so much of the literature relating to EAP comments on (see for

example Wong and Thang, 2008; Bannerjee and Wall, 2006; Watson Todd, 2006; and Chritison and Krankhe, 1986). More specifically, Christison and Krankhe (1986) explain that "in academic settings we certainly have a responsibility to teach the most accurate and appropriate language skills we can" (p. 76), and Watson Todd (2006) advocates establishing exactly "what the communicative requirements in [the new academic] environments are" (p. 149). Essentially, this chapter of the thesis will outline those aspects of language ability and, where appropriate, wider academic skills which are seen as key to determining how well students cope with the demands placed on them when they progress onto their main programmes of study (Bannerjee and Wall, 2006).

The chapter will consider each of the language skills individually, looking at listening, then reading, writing and finally speaking since this is primarily how they are tested. As will be seen, however, there is considerable crossover between the skills, and true separation is impossible. Owing to the nature of the discussion, there will be considerable cross-over with the previous chapter, and a lot of the details presented in Chapter 5 will be used to further support the discussions of the content and context validity.

6.2 What is academic listening?

Listening is a good starting point for the discussion of the assessments, since, as Read (2002) explains "so much of what they need to understand and learn is communicated through the oral medium" (p. 105), it is, therefore, imperative that the students have a high level of listening proficiency. Listening in academic contexts, however, is complex as the academic listening tasks "plac[e]… significant cognitive demands upon a student operating in their second language" (Taylor and Geranpayeh, 2011, p. 91). Therefore, this first section aims to establish the language knowledge and skills required to undertake the key listening tasks outlined in Chapter 5, that is to define the construct.

This has already been partially achieved in Chapter 5, and the key tasks which represent the construct of academic listening according to the focus group participants are:

- Listening in lectures to long stretches of discourse for the purpose of taking notes (both free and scaffolded)

- Listening in seminar discussions for the purposes of participating (listening to respond and co-construct meaning)

- Listening in presentations for the purposes of asking questions (listening to respond)

- Communicative listening in groupwork situations for the purpose of working with others to produce coursework.

In order to achieve these key tasks, it is clear that a certain level of language proficiency is required. Since the Common European framework of reference for languages (CEFR) is used for reporting the grades to students and the international student centre, a good starting point for establishing the predictive validity of listening is the CEFR descriptors. According to the details in the reference document, Table 10 below describes the overall listening comprehension at the level required for progression onto programmes.

Table 10 CEFR Common Reference Levels: Overall Listening Comprehension (Council of Europe, 2001, p. 66)

B2	Can understand spoken language, live or broadcast, on both familiar and unfamiliar topics normally encountered in personal, social, academic or vocational life. Only extreme background noise, inadequate discourse structure and/or idiomatic usage influences the ability to understand.
	Can understand the main topics of propositionally and linguistically complex speech on both concrete and abstract topics delivered in a standard dialect, including technical discussions in his/her field of specialisation.
	Can follow extended speech and complex lines of argument provided the topic is reasonably familiar, and the direction of the talk is sign-posted by explicit markers.

This, however, is just a starting point for defining what constitutes academic listening, and not all of the details reproduced above are necessarily helpful for academic listening situations.

As demonstrated in the literature review, many authors have outlined what constitutes listening in an academic context. Flowerdew and Miller (1995) for example, explain that academic listening skills include (but are not limited to):

- Note-taking

- Distinguishing relevant and irrelevant content

- Coping with long stretches of discourse

- Dealing with the multi-modal input usually encountered in a classroom situation

Bejar *et al.* (2000) define listening as "the process of retrieving an acoustic signal, which is then structured" (p. 2), explaining that listening comprehension consists of "a *listening* stage and a *response* stage" (*ibid.,* original emphasis). They go on to list four key listening purposes:

1. Listening for specific information
2. Listening for basic comprehension
3. Listening to learn, and
4. Listening to integrate information

<div align="right">Bejar et al. (2000, p. 10)</div>

Bannerjee and Wall (2006) provide an extensive list of listening skills which are also helpful in defining what constitutes academic listening, presenting a list of 'can do'

statements under three key headings: speed, note-taking and varieties of English. Included in their list is being able to respond quickly to information delivered at a variety of speeds, dealing with both native- and non-native-speaker accents, identifying main ideas and supporting points in lectures and seminars and making notes (Bannerjee and Wall, 2006).

These, however, are general descriptions of academic listening, and the purpose here is to establish whether the assessments are useful in assessing students' abilities to cope with the listening activities they will face at MUC. In order to do this, I will now consider the focus group discussions, starting with the points noted down during the sticky note activity which relate to listening tasks as listed in table 7 below.

Table 11 - Defining Academic Listening

English Development Team	Subject Specialists	Students
• Note-taking • Listening for main ideas/detail • Awareness of different teaching styles and expectations • Identifying lecture language • Knowledge of academic vocabulary – awareness of some of the subtleties • AWL items • Knowledge of subject-specific vocabulary • Academic register • Relative clauses • Complex noun phrases • Knowledge of hedging language • Clarification and comprehension	• Note-taking • Listening skills • Comprehension • Academic vocabulary	• Listening (lecture) • Making notes in the lecture

As can be seen from Table 7, the English development team made more comments about the specific language skills linked to academic listening tasks than the other two groups, as might be expected. This is an interesting contrast to the findings from

Christison and Krahnke (1986) in whose study students commented that "80% of their academically related language use was spent in reading and listening" (p. 70), thus implying that the students would have more to say about this area. This could, however, be due to the age of that study and the more general changes in education and assessment as were noted by several authors including Clegg and Bryan (2006), Boud and Falchikov (2007) and Norton (2007).

Moving now to the focus group discussions, academic listening was commented on during each of the focus groups. Starting with the students, during both groups, listening comprehension was discussed, albeit only briefly. In the first group, OM explained that lecturers "speak fast", and both students commented on the importance of note-taking in lectures. In the second student group, both participants also commented on the importance of listening to make notes. Coming to the subject specialists, in the first group, AL discussed the issues of dealing with longer stretches of discourse, commenting that "it's fine on a one-to-one basis, isn't it? In a lecture they can get confused" (AL, Subject Specialist Group One). In the second subject specialist group, AH explained that what was key for him in terms of listening was "the ability to listen and understand and comprehend" (AH, Subject Specialist Group Two).

For the English team, CL defined listening, explaining:

> So there's two lots of listening, isn't there? There's the receptive listening for note-taking, there's the receptive listening for responding in seminars. Surely the note-taking skills in lectures is more similar to note-taking from reading than it is from the listening skills?
>
> CL (English Development Centre)

This statement, along with the discussions in Chapter 5 demonstrate that students are also expected to undertake various "two way oral interactions" (Read, 2002, p. 105). Lynch (2011) and Taylor and Geranpayeh (2011) also comment on these two-way

interactions in which students are required to listen so that they can respond in some way. This reflects the report and discuss style of teaching which Mason (1995) explains is when the lecturers introduce topics in sessions, but these are used as cues for discussion points, with the lecturer intervening when and if necessary. Thus, the need to be able to discuss with, as well as listen to others to build on their points is an increasingly common approach (Ferris and Tagg, 1996a and 1996b), and a skill which needs assessing to establish if students will cope in their new environment. In fact, if Basturkmen (2002) is correct with her point that "discussion [is] interaction and talk and ideas [are] co-constructed" (p. 233), then the ability to listen to others and build on their points during a discussion is a further area related to academic listening which reflects the comments from CL, and which needs to be considered here.

This was not an area discussed at all by the subject specialists or the students, however the assessment audit shows that at level 6, seminars and debates make up 13.5% of the assessments and at level 7, they are16.6%. The ability to listen and then build on the ideas of others is, therefore, an important skill for university study and is certainly something which can be included in the assessments, although it may be better suited to inclusion in the speaking exams than the listening exams.

Based on the above comments, and bringing in the literature now, comprehension of audio material is the first aspect of defining the construct. Christison and Krankhe (1986), for example, comment that, in their study, "many subjects mentioned the initial difficulty in comprehending lectures" (p. 72). This comprehension is linked to two areas: knowledge of vocabulary and coping with longer stretches of discourse.

Starting with knowledge of vocabulary, Andrade (2010) explains that for many international students in her study "use of colloquialisms, unfamiliar cultural, political or historical allusions [and] unknown vocabulary" (p. 223) are key issues. In addition, Evans and Morrison (2011, p. 393) comment that often, students' problems with listening:

> stemmed mainly from their inability to understand key technical vocabulary, a problem their professors were often oblivious to and one that sometimes undermined the value of an entire lecture or seminar because puzzlement over terminology meant that they lost the thread of the explanation or discussion.

The 'oblivion' of subject specialist staff, however, was not reflected in this study, and KB in the second focus group explained that for her, it is important for students:

> to know the difference between certain words that we might use … in a lecture quite commonly, you might refer to a theory, then an author, then a reference, then a model, which is really the same thing. But what might, it might … it might be translated quite differently to, what's the difference between a model and theory? Well, nothing really.

KB (Subject Specialist Group Two)

This interesting point emphasises in particular the need of students to be familiar with the Academic Word List (Coxhead, 2011; Hyland and Tse, 2007; Coxhead, 2000), as was recommended by the English teaching staff, and reflects Evans and Morrison's (2011) comment that their participants often "lacked the requisite general and technical vocabulary to process [spoken] input" (p. 394). In the first subject specialist group, AL also touched on this point, commenting that it is important for students to be aware of "the nuances of how we use the English language" (AL, Subject Specialist Group One).

The Academic Word List (AWL) is a list of words first developed by Coxhead (1998) and presented for her MA thesis at the University of Wellington, New Zealand. The list is the "lexical items [which] occur frequently and uniformly across a wide range of academic material but [which] are not among the first 2,000 words of English"

(Coxhead, 2000, p. 218). In terms of coping in lectures, students' knowledge of the key language used in each part of a lecture helps them to follow the main points, which in turn supports their note-taking. For example, students need to be able to recognise language which indicates when the speaker is moving from one idea to the next, summarising a key point, emphasising an idea, giving an example or even coming to the end of a lecture. There are various criticisms of the AWL, including by Coxhead (2000) as well as Hyland and Tse (2007), however, and more up-to-date lists have been developed such as Gardner and Davies' (2014) Academic Vocabulary List and Ackermann and Chen's (2013) Academic Collocations List. The choice to use the AWL instead of these, despite the fact that it does not consider context, nor does it consider language items such as collocations was made because institutional practice insists on its use as part of assessment development.

The above points reflect the comments from Lewis and Reinders (2003) who recommend to students that they familiarise themselves with this key signposting language in order to make lecture listening more manageable. In fact, in the second student focus group, CY explained "we need to increasing more words, more vocabulary about the academic vocabulary" (CY, Student Group Two). NX responded to this, agreeing that having a good grasp of vocabulary was important. In the first student focus group, OM also commented "now I can understand really fast". On further questioning she explained that this related not only to the ability to deal with the speed of delivery, but also the fact that she was able to follow the main ideas of the lecture owing to her awareness of the signposting language and other verbal cues. DY discussed this area, albeit briefly, explaining that having a clear understanding of what the lecturers are talking about is partially aided by being aware of these verbal cues and useful signposting language. This further emphasises the need for students' knowledge

of and ability to use the AWL and thus the need to include items from the various sublists in the listening assessments.

In addition to vocabulary knowledge, a further point in defining the construct of academic listening is the length of discourse. The average lecture at MUC is two hours long, so being able to deal with longer stretches of discourse is key to coping with academic listening tasks. Participants in Evans and Morrison's (2011) study, who were very similar to the majority of pre-sessional students at MUC in that they came from Chinese-medium backgrounds, encountered difficulties with listening "because they were unaccustomed to listening to long stretches of discourse in English" (p. 394). This reflects the comment from AL quoted above.

To define the construct in the academic context on the basis of the comments from the focus groups and the skills listed above, then, listening for academic purposes involves any listening task which relates to academic life, including listening in classroom situations as well as the more general listening tasks related to studying such as communicating with lecturers in and out of class and coping with assessed tasks such as presentations and seminars. The context is, quite simply, academic. In terms of content, the assessments should include a range of language from the AWL, thus testing students' knowledge of this, as well as providing students with the opportunity to demonstrate their knowledge of key sign-posting language, used to organise longer stretches of discourse. Further, students need to be provided with opportunities to develop their collaborative listening skills, and I will now consider the assessments to evaluate how well they represent each of these points. In addition to this, I will evaluate the grading criteria and mark schemes to establish their reliability in terms of mark allocation and descriptors as well as how they operationalize the key skills noted above.

6.2.1 Vocabulary knowledge

Hyland and Tse (2007) explain that, in terms of the 'expected' breakdown of an academic text, on average 80% of it is made up of high frequency words, which are those words that make up the first 2000 words in English; AWL items should constitute between 8% and 10% of the text; the final 5% of the text would be technical vocabulary relating to the subject area.

Considering the listening assessments, the SEP listening exam includes a total of only 7% of the AWL, that is 40 out of the 570 AWL items. This is a very low coverage for an exam which purports to test *academic* listening. In terms of the language breakdown in each task, running the tasks through the vocabprofiler feature of Compleat Lexical Tutor[10] shows that these are as follows:

Table 12 - Vocabulary Profile SEP Listening Tasks

Vocabulary item	Task 1	Task 2	Task 3
First 1000 words	89%	77.8%	78.7%
Second 1000 words	3%	4.3%	10.3%
AWL	2.6%	9.3%	4.2%
Off-list/technical	5.1%	8.6%	6.8%

Taking each task in turn, the vocabulary profile for task one indicates very low coverage of the AWL, at approximately one quarter of the averages noted by Hyland and Tse (2007). This indicates that, in terms of vocabulary representation, this task is lacking in terms of the construct validity. Task two, a task in which students are required to note down information about key reading texts provided by a lecturer, is the most 'academic' in terms of vocabulary coverage, and conforms more closely to the averages.

---

[10] http://www.lextutor.ca/vp/eng/

Task three, the mini lecture, also has low AWL percentage. What can be inferred from this is that, because the construct of academic listening has been defined by the stakeholders as including familiarity with and ability to understand AWL items, the construct validity of this exam is compromised in that the predictions made on the basis of student performances on the tasks does not include how well they can cope with AWL items.

The IEP listening assessment fares slightly better in terms of the AWL item coverage, including 50 of the 570 items (8.8%). This is, however, still very low. The task breakdowns are as follows:

Table 13 - Vocabulary Profile IEP Listening Tasks

Vocabulary item	Task 1	Task 2
First 1000 words	83.8%	80.7%
Second 1000 words	2.6%	6.2%
AWL	7.8%	3.6%
Off-list/technical	5.8%	9.6%

Task one is the lecture task, and although the coverage of the first 2000 words is higher than the averages suggest it should be, the AWL coverage seems appropriate at just under the 8% average suggested by Hyland and Tse (2007). Task two, however, is questionable in terms of the vocabulary coverage, and with only 3.6% of the words used in the task coming from the AWL, it could be said that the task is lacking in terms of context and construct validity. This, in turn, impacts on the predictive validity of the task in terms of establishing how well students can deal with the AWL items in a lecture situation. Essentially, in terms of the vocabulary, this is an area which was seriously lacking in the exams in use between 2010 and 2014. By applying the evaluative framework, quite simply the assessments cannot be deemed useful in terms of their vocabulary inclusion.

6.2.2 Testing comprehension

Demonstrating comprehension of an audio text can be done in several ways as explained in the list provided by Bejar *et al.* (2000) above. They also explain that listening for specific information "requires comprehension of details of the text" (p. 10) and listening for basic comprehension "requires general comprehension of a text, such as comprehending all its main idea(s) and important details related to the idea" (*ibid.*), with listening to learn requiring a combination of both of these. Since the listening exams at MUC are testing whether students are able to make use of audio material for academic purposes, that is listening to learn, it seems sensible to assume that they will include tasks which assess students' ability to listen for specific details as well as to listen for basic comprehension of the main ideas and supporting details.

Looking at each of the exams in more detail, there are various opportunities for students to demonstrate their listening comprehension skills. Starting with the IEP, students are faced with the following tasks:

- Scaffolded lecture notes

- Sentence completion

- Table completion

- Comprehension questions to be answered using lecture notes

The SEP listening exam students have to complete the following tasks:

- Labelling a diagram

- Gap fill note-completion

- Noting down details about authors from lecture extracts

- Free note-taking

- Summarising a lecture from notes

- Using notes to answer comprehension questions

Comprehension is, to some extent, tested in all of these tasks, although more directly in some than in others. On face value, then, listening comprehension is tested extensively, however closer consideration of the mark schemes for these tasks is necessary, which I will do by looking first at the IEP exam and mark scheme, and then those for the SEP.

The first listening task faced by students during the IEP exam, as reproduced below, requires students to make notes using a scaffolded structure and then use these notes to answer comprehension questions at the end of the assessment. As discussed in Chapter 5, in terms of authenticity, this reflects more closely the reality of lecture listening at MUC.

Section One – How People Learn Best
Listen to the lecture and complete the framework. You will hear the lecture *twice*. Marks will be awarded for content (3) and abbreviations and symbols (2). You will use these notes to answer comprehension questions at the end of the listening exam.

Def. of multiple intelligences:

Traditional way to measure intelligence:

Reasons some people don't like this approach:

9 intelligences:

2 effects of multiple intelligences theory in classroom:
1)

2)

Example:

Issue with standardized test:

Figure 18 - IEP Listening Task 1

A potential area for improvement with this task, however, is the face validity, in that it could be made to look more authentic, with notes being presented in PowerPoint slides rather than simply presented in a word-processed document. This would enhance the face validity, and provide students with the opportunity to practice their note-taking in a way which more closely reflects reality at MUC. What does compromise the validity of this task is the fact that the students hear the information twice since this does not reflect the reality at all. At MUC, lectures are live and therefore only heard once, unless the students record the lectures themselves since lecture capture technology is not yet fully operational. Thus, giving the students the opportunity to hear the information a second time impacts on the predictive validity of this task since it gives a false impression of students' listening comprehension and ability to make clear notes on what they hear.

In terms of academic acculturation, as well as predicting how well students will be able to cope with live lecture listening tasks the lack of authenticity in this task essentially compromises the predictive validity. The task looks authentic, the construct, academic listening, is clearly represented, and the context, academic, is also evident, however, the predictive validity of the task, i.e. the inferences regarding how well students will cope with live lecture listening are clearly compromised and, therefore, this task needs revision. This need only be as simple as playing the task just once rather than twice so as to more closely reflect the reality of in-class listening tasks and presenting the class notes in a form which more closely represents what students will encounter in their lectures.

Considering the reliability of the mark scheme used for this task, there is no more information provided to the markers than that which is included on the student test paper. No example notes are provided for the markers; neither is there any clear breakdown for how the marks should be allocated. Additionally, only offering five marks out of a total of thirty for a text that makes up 50% of the exam could be questionable in terms of reliability, especially since this is the only task which is in some way reflective of the reality of lecture listening at MUC. With regard to the comprehension questions, there is less issue with the mark scheme reliability since students are presented with multiple choice questions, meaning that answers are either wrong or right (Figure 21 on page 192).

The second listening text in the IEP exam includes two direct comprehension tasks, as reproduced in Figures 19 and 20 below:

For questions 1 – 10, complete the ideas using **no more than two words or a number.**

1. By 2008, carbon dioxide emissions need to be ………………………………….. lower than in 1990.
2. Recycling ………………………………….. and reduces emissions from landfill sites and incineration plants.
3. For the survey, people were interviewed in …………………………………..
4. People say that one problem is a lack of ………………………………….. sites for household waste.
5. People sometimes put ………………………………….. and umbrellas in bottle banks.
6. Glass designed to be used for ………………………………….. cannot be recycled with other types of glass.
7. In the UK, ………………………………….. tones of glass is thrown away every year.
8. This amount consists of ………………………………….. bottles and jars.
9. In the UK, ………………………………….. tons of glass is recycled each year.
10. 50% of the glass collected is …………………………………..

Mark   /10

Figure 19 - IEP Listening Task Two

For questions 11 – 15, complete the table with information from the lecture.

Companies working with recycled materials		
Material	**Company**	**Product that the company manufactures**
Glass	CLF Aggregates	Material used for making **11**.................................
Paper	Martin's	Office stationery
	Papersave	**12** .... for use on farms
Plastic	Pacrite	**13**................................. for collecting waste
	Waterford	**14**.................................
	Johnson & Jones	**15**.................................

Mark   /5

Figure 20 - IEP Listening Task Three

Question Paper
Now use your notes from the lecture on multiple intelligences to answer the questions below. For questions 1 – 10 circle the correct option a – d.

1. What is another way to say multiple intelligences?
a. Multiple feelings
b. Multiple approaches
c. Multiple strengths
d. Multiple ideas

2. Which doesn't the speaker talk about?
a. The theory of multiple intelligences
b. Nine kinds of intelligence
c. How the multiple intelligence theory affects a classroom
d. Why some people oppose the theory of multiple intelligences

3. Why don't some people like intelligence tests?
a. Because the average score is 100
b. Because they don't accurately measure all strengths
c. Because an above average score is 130
d. Because they're often very long

4. Which is true of the theory of multiple of intelligences?
a. Everyone has the same strengths
b. Everyone has all nine intelligences
c. Everyone has some intelligences that are stronger than others
d. Both b and c

5. What does the speaker believe about all students?
a. They need to be allowed to demonstrate their understanding in different ways
b. They have high IQs
c. They have strong verbal intelligence
d. They would benefit from a tree-drawing exercise

6. What are the two characteristics of someone with verbal intelligence?
a. Has a good sense of direction and can identify shapes
b. Communicates well in words and uses language skilfully
c. Has a strong connection to nature and one's place in it
d. Has an ability to analyse visual images and charts

7. Someone who uses logic to solve problems is using which intelligence?
a. Artistic
b. Kinaesthetic
c. Mathematical
d. Spatial

8. Alex is a great leader and works well with others. Which intelligence does this show?
a. Interpersonal
b. Intrapersonal
c. Artistic
d. Kinaesthetic

9. What is the speaker's attitude towards allowing students to write poems or songs?
a. It's unfair to all of the students
b. It's less accurate than giving a written test
c. It makes it difficult to really assess students
d. It's reasonable based on multiple intelligences theory

10. How do you think the lecturer probably feels about standardised tests?
a. She finds them useful
b. She thinks they don't measure all types of intelligences
c. She believes they benefit most students
d. She thinks they discriminate against all students

/10

Figure 21 - IEP Listening Question Paper

IEP Listening tasks two and three assess students' ability to listen for specific details, as outlined by the literature and the English teaching team as key for assessing listening ability. The mark schemes for this are simple, with just one mark per answer, no matter how much information is required to complete the sentence or box. This could lead to questions about the reliability, since students are awarded one mark for "green" and "roads" (questions 10 and 11), but also one mark for "reduces greenhouse gas" and "soil conditioners" (questions 2 and 12).

With regards to testing comprehension, then, the IEP exam is useful, since students have several opportunities to demonstrate this skill. Despite this, the tasks could be said to have low context validity although they represent the construct well.

Coming now to the SEP, it consists of five parts, all of which offer students the opportunity to demonstrate their listening comprehension skills in different contexts, thus in terms of the construct, the test has high construct representation. Where this exam has potential issues is with the content of some of the tasks in terms of topic representation. Part one, for example, is focussed on students joining a library; the 'academic' value of this is questionable, hence the context is not well represented in this first part.

In contrast, part two (see Figure 22 on the following page) could be considered to have high context validity, in that it reflects the academic context, replicating an activity which students might be expected to do on progression, since it requests that they note down author names and details about their publications. Additionally, in terms of content validity, the task presents students with a variety of key functional language. In terms of face validity, however, this task is questionable; it is unlikely that lecturers

would expect students to note down these kinds of details without providing them with the spelling of the authors' names. This is not done in the exam, and it seems unlikely that students would be able to spell, accurately (as is required by the mark scheme), these particular names.

> **Part Two – Listening section 2**
>
> Names must be correctly spelt
>
> 11. Name: Neil Leiper [1 mark]
> Extra information: 1 mark each for any two of: book / tourism management / one of the core texts
>
> 12. Name: McLeod [1 mark]
> Extra information: 1 mark each for any two of: article / 2006 / Cultural Commodification and Tourism: A Very Special Relationship / Tourism Culture and Communication Volume 6
>
> 13. Name: Walter Pauk [1 mark]
> Extra information: 1 mark each for any one of: How to Study at College / 2004 / 8th Edition
>
> 14. Name: thefreedictionary.com [1 mark]
> Extra information: Web source [1 mark]

Figure 22 - SEP Listening Exam Part 2 Mark Scheme

Parts three to five (see Figures 23 to 24 on the following pages) of the exam all focus on one text: a lecture about owning a successful restaurant. This is a free note-taking task, which reflects well the construct of academic listening, and provides students with two opportunities to demonstrate their comprehension: once in a summary task where they have to summarise the key points of the lecture, and again in a series of short-answer comprehension questions which students must answer using their notes. This provides high levels of face, context and construct validity, thus the task appears useful. However, on considering the mark scheme, the reliability of these tasks could be questioned.

> **Part Three – Listening Section 3**
> No marks awarded for the notes – only mark the summary
> Summary marks /15:
> - Key points noted down – 5
> - Spelling and grammar – 5
> - Organisation – 5

Figure 23 - SEP Listening Exam Part 3 Mark Scheme

The reliability of this is clearly questionable, since no example summary has been

provided, and no breakdown for each area is included in the mark scheme.

Additionally, despite no mention of this in the task instructions, the summary is being

marked for spelling and grammar, which indicates a lack of validity in terms of

construct irrelevant factors (Aryadoust, 2013). The application of this mark scheme is

also likely to be subject to considerable rater bias. Additionally, the marks awarded are

not likely to be free from errors of measurement as required by Weir (2005) since there

is very little information provided for the raters in the first place. If we take these points

into consideration, along with the discussion of the task itself, the inferences made

about students' abilities to cope with lecture listening on the basis of the marks awarded

for this task are likely to be invalid, unreliable, and ultimately not useful.

Considering now the comprehension questions (see Figure 24), there are various issues

in terms of the reliability. The main issue is how the marks are awarded. Each question

is given just one mark, yet there is a considerable discrepancy in the amount of

information required for each question. Questions one and four and five, for example,

require considerably more details than does question three, yet they all have the same

weighting. Further, the students are working from the notes they took whilst listening

to the recording, so it is possible that they could be penalised twice for not noting down

the key points. Despite these obvious issues, the comprehension questions are clear,

and are easy to mark. Further, using the notes and the summary to answer

comprehension questions reflects the reality of academic life, in which students are often required to use their class notes to revise for exams or inform their essays and reports. Hence, in terms of the evaluative framework, the authenticity of this particular task aids academic acculturation and further enhances the predictive validity of the assessment, ensuring that the students' ability to deal with this kind of task is assessed.

```
Part Five – Listening Section 4.
_____
(1) What are the two main concepts that the lecturer wants you to remember from
the lecture?
a. Many factors lead to customer satisfaction [ ½ ]
b. A successful dining experience is built on contrast [ ½ ]

(2) What are the two main groups that restaurant staff are divided into?
a. Front of house [ ½ ]
b. Back of house [ ½ ]

(3) If someone wanted to open a restaurant because they love food, what advice
would the lecturer give them?
Learn the business [1]

(4) Would this lecture be relevant to a manager of a MacDonald's restaurant?
Explain your answer.
No because the lecture is about independent restaurants not chains [1]

(5) What piece of information in the lecture does the lecturer thing is the most
surprising for the listeners?
More than half of all restaurants are independent [1]

                                                                          /5
_____
```

Figure 24 - SEP Listening Task 4 Comprehension Questions and Answers

6.2.3 Collaborative listening tasks

As explained above, the one-way lecture-style listening constitutes just one type of academic listening task, and students also need to be able to listen and respond in seminar, presentation and groupwork situations (Kingston and Forland, 2008; Kember, 2000; Ferris and Tagg, 1996a). Looking at the assessments, it is the speaking tasks where the communicative listening is included, and the table below lists the key

instructions relating to how the students need to be able to listen to others and build on their ideas.

Table 14 - Speaking Tasks Linked to Communicative Listening

English assessment	Relevant task instruction
SEP – Seminar	Show the ability to respect turn-taking rights and to interact with colleagues, responding to and extending points raised.
IEP – Digital portfolio task 1	Remember to make relevant comments in response to your partner
IEP – Digital portfolio task 2	Remember to make relevant comments in response to your partner
IEP – Digital portfolio task 3	Remember to make relevant comments in response to your group members
IEP – Seminar	Show the ability to respect turn-taking rights and to interact with colleagues, responding to and extending points raised.

It is clear, then, that the communicative listening skills required to co-construct meaning, and needed according to the comments from the focus groups, are included in the speaking assessments, thus the speaking assessments are usefully covering the elements of collaborative listening. Certainly, these instructions reflect Basturkmen's (2002) comments emphasising that meaning in academic situations is often co-constructed and therefore relies on the students listening carefully to others in order to do this. Hence, the inferences made about students' abilities on the basis of these tasks are valid in that the authentic approach to the speaking tasks suitably incorporates this particular element of the construct of academic listening with appropriate content. But this is just the first part of this construct discussion, and the grading criteria also need to be investigated here.

Considering how the grading criteria operationalize these skills the table below contains the descriptors from the grading criteria. Both programmes make use of the same grading criteria when marking the speaking assessments.

Table 15 - Speaking Grading Criteria Operationalization of Communicative Listening

Percentage grade	Descriptor
85 – 100	*No descriptor provided*
75 – 84	Very good effort made to pay attention to other students
60 – 74	Good effort made to pay attention other students
50 – 59 (borderline fail)	Some effort to pay attention to other students / respond to questions
40 – 49 (fail)	Some effort to pay attention to other students / respond to questions
35 – 39 (fail)	Fails to pay attention to other students / respond to questions
Below 35 (fail)	No attempt made to contribute / interact with other students

Although the grading criteria do attempt to include reference to how well students listen to others, there is the obvious issue with the fact that this is not included in the 85% - 100% band. Additionally, there is the fact that the 40% - 49% and 50% - 59% bands contain the same descriptors, making it impossible for markers to score this area reliably. The reliability of the grading criteria, then, are instantly called into question and this indicates the need for amendments to these.

In addition to the lack of consistency across the bands, there is also the issue that in terms of communicative listening, the criteria only comment on listening to and paying attention to others. The additional communicative listening tasks as outlined above, namely respecting turn-taking rights and co-constructing meaning, are not included in these criteria, further undermining their reliability.

6.2.4 Summary comments

Having applied the evaluative framework, the discussion above and in Chapter 5 indicates that the listening assessments used on both the IEP and the SEP require a great deal of reworking on all levels. The predictions which were made on the basis of the marks awarded for the assessments cannot be considered to be valid since, although the tasks included in the exams were measuring the broad construct of listening, the academic context was lacking. The face validity of the tasks could also be improved with the addition of scaffolded notes which are completed on PowerPoint slides. Certainly, the non-lecture-based tasks are not useful tests of listening for academic purposes. What has emerged from this discussion, then, is that amendments are required, with new scripts which incorporate a wider spread of the AWL items, and which are perhaps longer stretches of discourse. Additionally, a more direct test of signposting language that is used in lectures in the form of a cloze test could be included in the exam. Finally, with both sets of assessment being used to make the same decisions, and at the end of programmes with the same overall learning outcomes, the assessments need to be more similar in nature, with the strongest elements of each being combined.

In terms of the reliability of the mark schemes, there is insufficient information included for markers in terms of what would constitute 'adequate' notes. Further, information such as how students could achieve full marks for the notes in terms of their use of symbols and abbreviations, or an example summary would increase the reliability of the marks awarded to the students. Although there will always be an element of subjectivity in the marks awarded on these assessments, examples and more detailed information for markers would reduce the impact of this. Thus, the results of the

evaluation indicate that improvements are also required with the mark schemes to improve their reliability, and thus enhance the overall usefulness of the assessments.

6.3 What is academic reading?

A great deal of the literature regarding academic skills explains that good reading skills are vital for students to cope with their academic life. Evans and Morrison (2011), for example, explain that "reading is crucial in the process of disciplinary socialization" (p. 392), and Andrade (2006) highlights the "heavy reading load" (p. 139) which students face when they study in English-medium universities. With this in mind, Spector-Cohen *et al.* (2001) state that "the ultimate goal of the EAP reading course is to enable students to participate in the discourse communities of their future fields" (p. 376). Andrade (2010), however, explains that, as would be expected, "NNESs [non-native English speakers] tend to read more slowly … and may be unused to the discourse patterns used in written text" (p. 223). Spector-Cohen *et al.* (2001) thus emphasise that the aim of EAP reading courses, and therefore the assessments used on them, is to "equip [students] with a repertoire of strategies to become independent readers" (p. 375), and in order to do this, it is vital that the English team are able to establish what this repertoire of strategies should entail. That is the purpose of this section, and Table 16 on the following page outlines the key reading skills as noted down by the focus group participants.

Table 16 - Defining Academic Reading

English Development Team	Subject Specialists	Students
• Note-taking • Skim/scan/detail • Identifying main ideas in text • The ability to take notes for a particular topic from extended reading materials • AWL items • Knowledge of subject-specific technical vocabulary • Knowledge of academic vocabulary – awareness of some of the subtleties • Extended noun-clause recognition • Relative clauses • Identifying parts of speech (and their purpose) • Complex noun phrases • Effective reading skills • An ability to scan through extended reading material for key words related to a topic of interest	• Reading • Note-taking • Sketch-noting • Questioning approach • General understanding • Comprehension • Reading and analysing articles • Academic vocabulary • To read texts not only factual but fictional as well • Quick/speed reading • Understanding reading texts • Pre-reading suggested materials	• Explore a case related to lectures • The abilities to find core-sentences/key words • Read course introduction • Read news for seeking information

As is clear from the above table, there are several fundamental areas of study which are related to students' academic reading ability. The most important arising from the focus groups were linked to: general reading proficiency, the ability to take effective notes from reading, which links to reading for a specific purpose, and critical reading skills. To define the construct, then, academic reading involves reading a variety of long and complex texts about various subjects related to studies in order to learn about a particular subject or for the purpose of using them to support further academic work, both written and spoken. These areas reflect the specific reading purposes as highlighted by Enright *et al.* (2000), namely "reading to find information, reading for basic comprehension, reading to learn and reading to integrate information across multiple texts" (p. 1). Since note-taking, reading for a specific purpose and critical reading have already been discussed in some detail in chapter five, this section will

consider just the first two areas: reading to find information and for basic comprehension and vocabulary knowledge.

6.3.1 Reading to find information and basic comprehension

Starting with general reading proficiency, that is the reading to find information or reading for basic comprehension, as Bygrave *et al.* (2014) explain "in Westernized cultures, knowledge is acquired through synthesizing and reconstructing the words of experts" (p. 209), and this implies not only a strong emphasis on vocabulary knowledge, but also knowing how to deal with complex written structures such as the complex noun phrases or defining relative clauses as discussed by the English teaching team. This supports Evans and Morrison (2011) who explain that "processing such complex and dense texts can be extremely challenging and time-consuming" (p. 120). Without good vocabulary and grammar knowledge, comprehension of the sources would be difficult, which would impact on the ability to effectively paraphrase and synthesise the work of others. Table 16 on the previous page demonstrates the English team's and subject specialists' acknowledgement of this. In the discussions with these different groups, CW in the first subject specialist focus group commented that:

> if [students] read then they get the language to be able to use in the essays that makes their essays better, gives them a better understanding. It's kind of a holistic thing really.
>
> CW (Subject Specialist Group One)

In the second subject specialist group, AH also commented on reading comprehension, explaining that a key area of reading for him is "basic, fundamental reading skills. And to have some comprehension of what they're reading" (AH, Subject Specialist Group Two), a point also picked up by KB who highlighted that "being able to understand [what they read] and then being able to communicate that" (KB, Subject Specialist Group Two) are important for coping with academic studies.

The reading assessments used on both programmes, then, need to include reading comprehension tasks, which they both do. Starting with the SEP, there are two tasks which directly test reading comprehension. The first is task two, which is reproduced below.

Task 2:
Read the article 'Unlikely Boomtowns' and answer the following questions. You must write the answers IN YOUR OWN WORDS. Do not copy sentences from the article.

(1) Describe the changes in growth for cities with more than 10 million inhabitants before 2005.

(2) What size of city do most people live in?

(3) What is the main point in the article?

(4) Why is property in megacities so expensive, according to the article?

(5) What is a 'dormitory town'?

(6) Why have commute times in many countries increased?

/6

Figure 25 - SEP Reading Exam Task 2

This task tests scanning, a vital, basic, reading skill which requires students to work quickly through a text to find certain details to help them answer a question; something which will aid them when they progress. It also tests their general comprehension of the information related to the comments in the questions. Thus, in terms of usefulness, there is a clear match between the construct being tested and the skills required by students when they progress. Additionally, the text itself 'looks' academic since it includes references. What is a concern, however, is the mark scheme, reproduced in Figure 26 on the following page:

Figure 26 - SEP Reading Exam Task 2 Mark Scheme

Although on first viewing, the mark scheme here looks reliable, there are several points of contention. Firstly, the students are required to paraphrase the answers from the text, instantly introducing an element of subjectivity on the part of the marker since marks are only awarded for a 'correct paraphrase'. What constitutes a 'correct paraphrase' is highly subjective, as is how much of the original language needs to be changed for the answer to be considered a paraphrase. Additionally, what exactly is meant by 'an approximation of these' is also unclear. With no clear instructions or examples of a 'correct paraphrase' or what constitutes an acceptable 'approximation', the students might not receive marks for their answers. Additionally, there is just one mark awarded for each question, regardless of the amount of detail required in the answer. When considering, for example, questions two and four, there is a considerable difference in the amount of information required to answer the questions, and, therefore, the amount of paraphrasing required. The final issue with this mark scheme is that students are not

awarded any marks for copying from the text, even though they may have identified the correct information to answer the question. This reduces the construct validity of the task since students are being tested on their paraphrasing skills rather than just their reading skills. Although it could be said that paraphrasing is a key academic skill, hence this enhances both the context validity and authenticity of the task, the fact that this is a comprehension task means that students should, ideally, be rewarded for finding information which answers the question, even if they are not able to put the ideas into their own words.

The second task in the SEP reading exam which assesses reading comprehension is task four (reproduced on the following page in Figure 27), which is based on the second reading passage. This time the students are required to match names to statements from the text. This task tests both comprehension and scanning skills (amongst other things); two basic but important reading skills, as evidenced by the comments from the focus groups and, in particular, the English team sticky note activity. It is also an effective way to test vocabulary knowledge, as will be explained in section 6.3.2 below. Thus, the validity argument being made here with regard to the construct and context validity is strong since the task reflects key academic reading skills. Although the task lacks authenticity, in that students are rarely expected to answer comprehension questions related to texts they have read, this is offset by the high levels of construct and context validity.

Stephen Boyden	Gordan Orians	Roger Ulrich
Rachel Kaplan	Stephen Kaplan	Terry Hartig
Francis Kuo	Walter Kroner	Melvin Konner

1. An individual's mental well-being can be improved by giving more control to them.

2. The environment that an animal evolves in will be the healthiest environment for them to live in.

3. Looking at a forest will make you happier.

4. Violence in cities might be reduced by having more plants in the street.

5. Perhaps humans like to meet together to eat because of evolutionary forces which favoured people who met together in the evening to talk about the day and think about the next day.

6. Having a natural view from your window will make you happier.

7. Humans might prefer wide, open spaces because of evolutionary forces.

/7

Figure 27 - SEP Reading Exam Task 4

In terms of marking, this is a simple task to mark, with answers simply right or wrong, hence there is no possibility of rater bias. Therefore, with regard to its usefulness, this particular task could be considered useful for all stakeholders. The English team are able to make valid and reliable predictions about how well the students will deal with this type of task when they progress; the students are given the opportunity to demonstrate their academic reading proficiency and are also provided with examples of the academic reading tasks they will have to be able to do to cope with their course reading; and finally it is useful for the subject specialist lecturers since they can be confident that the students have been given the chance to demonstrate their ability in a suitably academic context.

The final task in the SEP reading exam which tests reading comprehension is task 6 (see Figure 28).

Figure 28 - SEP Reading Exam Task 6

Here, students are required to predict in which paragraphs of an article they might find certain information; this task tests both comprehension and predicting skills. Additionally, this task tests comprehension of the key words and ideas, demonstrating content validity. It is also reliable in terms of marking since the answers are either right or wrong, thus the mark allocation requires no subjective interpretation.

Moving to the IEP reading exam, the majority of the tasks incorporate some element of reading comprehension. This skill is most directly tested in task C, however not all of the activities relate to comprehension. Those relevant sections are reproduced below in Figure 29.

Task C:
Read this extract on the Impact of Globalisation on Equality and answer the questions 1 – 15 below.

3. Explain the significance of the following numbers within in the text. Try to use your own words.

30% - 40% _____

_____

1960 _____

_____

1.2 billion _____

_____

358 _____

_____

2.3% - 1.4% _____

_____

[5 marks]

Figure 29 - IEP Reading Exam Task C Comprehension Task Extract

As with the comprehension tasks in the SEP reading exam, this task requires students to scan the text to find the required number and then read intensively to be able to answer the questions. This is an effective test of the key skills of scanning and comprehension.

The mark scheme (Figure 30 on the following page) is somewhat problematic, however, since it is subjective, and requires construct irrelevant factors for students to be awarded full marks. However, although this mark scheme has the same issues as that used for the SEP in terms of subjectivity and the potential impact of rater bias, it does at least require markers to acknowledge students' ability to locate key information, even if they cannot necessarily put that information into their own words. This approach enhances

the construct validity of the task since students are rewarded for being able to read and understand the necessary details.

Task C:
Read this extract on the Impact of Globalisation on Equality and answer the questions 1 – 15 below.

_____

3. Explain the significance of the following numbers within in the text. Try to use your own words.

30% - 40% **the percentage of the world population that has benefitted from globalisation**

1960 **the year when the income ratio between the richest and the poorest was 30:1**

1.2 billion **the number of people living on less than $1 per day**

358 **the number of people who own as much wealth as 2.5 billion other people have**

2.3% - 1.4% **the drop in the global income of the poorest 20% of the world's population**

These are paraphrases; equivalents of this are acceptable. Copying from the text – ½ marks

[5 marks]

Figure 30 - IEP Reading Exam Task C Mark Scheme

As well as testing students' comprehension directly, both reading exams also test comprehension in the note-taking tasks. In the SEP exam, students are asked to "write notes to record the main trends identified by the article, and the reasons suggested for these trends" (SEP reading exam, 2010), which requires students to be able to understand the key points of the text they read, with marks awarded for accuracy of content. Additionally, the IEP exam has a similar task which asks students to "make notes on the key points discussed" (IEP reading exam, 2010); again, students are awarded marks for content, since they are required to be able to understand the key points in order to be able to note them down. The mark schemes for the tasks might be considered problematic in terms of reliability, with both simply stating that students should be awarded three and four marks respectively, with no mention of how to differentiate between zero, one or two, for example.

So far, then, both of the reading exams could be said to be valid from the construct, face and context point of view, and hence the predictive validity is relatively certain with regards to how well the assessments test reading comprehension and skimming and scanning skills. However, when considering how the tasks listed above are marked, there are clearly reliability issues in terms of rater bias and subjectivity. In order for the exams to be considered useful, then, greater emphasis needs to be placed on the students' ability to locate information to answer the questions rather than their ability to paraphrase these ideas. Paraphrasing skills can be more usefully and authentically tested in other ways as will be discussed in section 6.4 below. Additionally, the mark schemes for the note-taking tasks could be more explicit in terms of detailing how raters should award, say, one mark instead of two.

6.3.2 Vocabulary knowledge

Linked to general comprehension is students' knowledge of and ability to manipulate key language, a key factor in terms of the construct of academic reading. With regards to comprehension and vocabulary in the literature, Evans and Morrison (2011, p. 392) comment that their participants

> experienced considerable difficulties in adapting to the demands of university-level reading … [many of which] stemmed from their lack of a rich technical vocabulary, which prevented them from quickly and fully comprehending unfamiliar disciplinary genres.

They (*ibid.)* add that for some students "their knowledge of 'general' academic vocabulary … was also incomplete" (p. 392). Andrade (2010) also discusses this area, explaining that for many international students "vocabulary knowledge may not be broad enough for complete comprehension" (p. 223) of written discourse. Similar comments also came from Ostler's (1980) study, which calls for EAP programmes to "develop strategies in their students for understanding the rhetoric of various types of writing, such as those found in academic journals and papers" (p. 492). He adds that

"students can often understand the words and the sentences, yet not comprehend the total meaning of technical discourse" (Ostler, 1980, p. 492).

Comments from the students reflect these points, and CY in particular explained

> I was experiencing, er, there's too much article, too academic. It's quite difficult to understand to be honest, and, I mean, for totally understanding because some words is so unfamiliar, I never see. I actually never see them before.

CY (Student Group Two)

In the first student focus group, both participants also mentioned the importance of vocabulary knowledge in relation to reading. This further supports Evans and Morrison's (2011) comments above which relate in particular to students from China. Since the majority of students on the pre-sessional programmes at MUC are from China, Taiwan and Hong Kong, it can be assumed that the same applies to these students.

During the English team discussion, vocabulary knowledge in relation to reading was, as would be expected, afforded a great deal of importance, with particular mention made of the AWL by three of the participants during the sticky note activity. In the discussion, LM highlighted the importance for students to have an understanding of "academic terminology", and CL explained that he noted down 'knowledge of vocabulary' in relation to the reading. SD also mentioned vocabulary, and for him it is the AWL items that are the most important. These points echo the comments from the literature above.

There are several opportunities provided for the students to demonstrate their vocabulary knowledge in both reading exams. Starting with general statistics, the SEP

reading assessment had the highest coverage of the AWL, including 111 of the 570 items, with the IEP reading exam including 72 items, that is 12.6%. Looking at the individual texts and tasks in more detail, the table below shows the percentage breakdowns of the separate tasks:

Table 17 - Vocabulary Profile SEP Reading Texts

Vocabulary item	Passage 1	Passage 2	Passage 3	Passage 4	Overall
First 1000 words	76.2%	71.1%	58.3%	75.7%	70.3%
Second 1000 words	2.7%	6.3%	17.3%	1.5%	7.0%
AWL	6.1%	11.7%	7.9%	16.9%	10.7%
Off-list/technical	14.5%	10.9%	16.5%	5.9%	12.0%

Drawing on the comments from Hyland and Tse (2007), passage one has quite a low percentage of AWL words, yet the off-list words are considerably higher. This could, however, be due to the high number of names included in the text to be used for a referencing task. Despite the low overall percentage of AWL making up the text, there are 29 of the AWL items in this text. Passage two, which is the longest passage in the exam at 950 words, includes 59 of the AWL items, and in general the percentages are similar to those averages cited by Hyland and Tse (2007). The same, however, cannot be said of passage three, which is an abstract from an academic article, but which contains a high number of off-list words. Unlike passage one, where this higher figure can be attributed to the names included in the text, passage three does not contain any references; what could explain the high number of off-list words, however, is the fact that the article is focussed on a very specific topic – Chinese and Taiwanese hotel website design. In terms of AWL coverage, passage four is particularly interesting since it is topic sentences only, and contains only 149 words. Of these words, 16.9% of them are AWL words, and the text contains 19 items from the AWL. Essentially, in general this exam exposes the students to, and thus tests their knowledge of, quite a broad range of AWL items. Therefore, in terms of the construct requirements of the

participants, supported by the literature, the predictions made on the basis of this exam in terms of students' knowledge of academic vocabulary could be said to be valid. This needs further investigation, however, since there are many ways in which students can demonstrate their vocabulary knowledge, as will be seen in the rest of this section.

Moving now to the IEP reading assessment, the percentages are below:

Table 18 - Vocabulary Profile IEP Reading Texts

Vocabulary item	Abstract 1	Abstract 2	Abstract 3	Task C	Overall
First 1000 words	65.4%	63.5%	67.0%	78.9%	75.3%
Second 1000 words	12.8%	2.1%	5.2%	3.3%	4.8%
AWL	17.9%	28.1%	23.5%	9.0%	12.6%
Off-list/technical	3.8%	6.3%	4.4%	8.8%	7.3%

When comparing these averages again to Hyland and Tse (2007), the abstracts used for tasks A and B have a particularly high percentage of AWL items, especially in abstracts two and three. Looking at each abstract in more detail, abstract one includes 11 of the AWL items, abstract two includes 13 and abstract three 21. Across the three, students are exposed to 42 of the AWL items in total, a large number across such short texts. In approximate figures, the text used for the rest of the exam contains the expected number of AWL words, including 27 separate AWL items in total. In contrast to the SEP exam, the IEP has a lower overall coverage of the AWL, so in terms of predicting how well students know this language, it could be said that the SEP reading exam is a more valid assessment than that used on the IEP. However, this is, as explained above, quite a crude measure of language coverage, and testing vocabulary needs more than just a list of statistics relating to vocabulary coverage.

As well as investigating the general AWL coverage of the exams, however, there are other factors to consider when addressing the validity of the assessments in terms of

construct representation; it is also necessary to look at how the assessments actually test students' knowledge of, and proficiency with, using vocabulary. As mentioned above, in the SEP exam, task four tests vocabulary knowledge in that the statements in the questions are paraphrases of the information in the text, meaning that students' ability to find synonyms is being tested, as is evidenced in Figure 29 below, which shows the statements from the questions and the corresponding section of the reading text. Looking at the contents, the students' vocabulary knowledge is tested quite extensively in this task, with complex structures from the original being paraphrased and summarised quite considerably in the questions. In terms of 'academic' vocabulary, the Compleat Lexical Tutor breakdown of the text as listed above in Table 17 shows that 11.7% of the text comes from the AWL, with a total of 59 AWL items included.

Question statement	Corresponding original text
1. An individual's mental well-being can be improved by giving more control to them.	...personal control leads to significant increases in comfort and morale.
2. The environment that an animal evolves in will be the healthiest environment for them to live in.	...the optimum healthy environment ... 'the conditions which tend to promote or permit an animal optimal psychological, mental, and social performance in its natural or "evolutionary" environment'.
3. Looking at a forest will make you happier.	...workers with window views of trees had a more positive outlook on life...
4. Violence in cities might be reduced by having more plants in the street.	...outdoor nature buffers aggression in urban high-rise settings...
5. Perhaps humans like to meet together to eat because of evolutionary forces which favoured people who met together in the evening to talk about the day and think about the next day.	The ritual of sitting round a fire on the savannah or in a cave telling stories of the day's stories of the day's events and planning for tomorrow may be an ancient carryover from *Homo sapiens'* hunting and gathering days.
6. Having a natural view from your window will make you happier.	...passive viewing of nature through windows promotes positive moods.
7. Humans might prefer wide, open spaces because of evolutionary forces.	...humans are psychologically adapted to and prefer landscape features that characterized the African plain or savannah, the presumed site of human evolution.

Figure 31 - SEP Reading Exam Task 4 Extracts

Yet this is not the only task which tests students' vocabulary knowledge in the SEP reading exam thoroughly, since Task 3 of the exam (see Figure 30) requires students to

make notes, using their own words. This task entails not only summarising, but also extensive vocabulary knowledge for the ideas to be put into different words.

Task 3:
Read the article carefully, and write notes about the article to record the main trends identified by the article, and the reasons suggested for those trends. Remember to record details for sources if necessary. You give your notes clear organisation, and you must not write whole sentences. **DO NOT COPY WHOLE SENTENCES DIRECTLY FROM THE ARTICLE.**

/10

Figure 32 - SEP Reading Exam Task 3 Instructions

This task clearly requires the students to understand the key ideas, and to have the vocabulary knowledge to be able to put these into their own words. A similar task is included in the IEP reading exam:

Task D:
Re-read the extract on the impact of globalization on equality and make notes on the key points discussed which relate to the essay question in Task A. Remember to your notes **using your own words**.
You will need to make use of your notes to write a summary in Task E.
Marks are awarded for:
- Content (up to 4 marks)
- Paraphrasing (up to 4 marks)
- Layout (up to 4 marks)

Figure 33 - IEP Reading Exam Task D Instructions

As can be seen, both of these tasks instruct the students to use their own words, which offers the students the opportunity to demonstrate their vocabulary knowledge. What is problematic again is the subjective nature of the mark schemes, in particular with the SEP with no model notes provided for the markers. Although the IEP exam does provide markers with a breakdown of how the marks should be allocated for each area (Figure 33), there is no indication as to what should be awarded full marks, what should get three marks and so on.

The IEP exam also provides students with one final opportunity to demonstrate their vocabulary knowledge. In order to support the summary task in the exam, the students are required to match words from the article they read to suitable synonyms (Figure 34). Both assessments represent the academic reading construct effectively, and provide the students with a variety of opportunities to demonstrate their vocabulary knowledge.

Match the words from the extract (4 – 10) with the synonyms/definitions and write the answers (a-g) in the spaces provided below	
4. imported	a) undeniable
5. shoddy	b) ignored
6. expert	c) brought in from another country
7. chasm	d) differing
8. marginalized	e) specialist
9. contradictory	f) inferior
10 incontrovertible	g) large gap

4 = ____   5 = ____   6 = ____   7 = ____   8 = ____   9 = ____   10 = ____

Figure 34 - IEP Reading Exam Task C Extract

6.3.3 Summary comments

The discussions above, along with comments from Chapter 5 again indicate that considerable work is required in terms of improving these assessments in all areas. Certainly, both assessments have strengths, not least of which is their inclusion of tasks which assess vocabulary knowledge; unfortunately, the AWL coverage in the texts chosen is, in general, lower than the recommendations from Hyland and Tse (2007), thus bringing the content validity into question. There are some useful tasks included in the assessments, but generally, in terms of context validity, greater attention needs to be paid to the academic reading tasks which students will be faced with. Tasks which mirror these more closely could then be included in the exams to enhance their overall predictive validity.

Concerning the reliability of the mark schemes, this is questionable since, as with the listening assessments, they both lack examples and clear explanation of how the mark schemes should be applied. By offering this type of information for the markers, although subjective marking is impossible to eliminate entirely, its impact could be reduced.

In terms of the construct representation, then, the assessments do test the key reading skills as identified above, that is reading to find information, reading for comprehension and testing vocabulary knowledge. However, in order for the assessments to be considered useful assessments of academic reading skills, work is required in the areas of reliability and context validity.

6.4 What is academic writing?

Academic writing, as the literature demonstrates, is a complex skill with which students need to quickly become familiar if they are to be able to cope with their HE level studies (see for example Donohue and Erling, 2012; Andrade, 2010; Wong and Thang, 2008; Harmer, 2007; Andrade, 2006), and this is the same at MUC. In fact, AH in subject specialist group two stated "I'm thinking about their English writing as a priority". In terms of the sticky-note activities, Table 19 on the following page outlines what the research participants highlighted relating to academic writing. Unsurprisingly, the English team made more reference to the technical language side of academic writing, which reflects Donohue and Erling's (2012) comments that "a subject lecturer is unlikely to pay primary attention to *how* the language is used" (p. 215).

Table 19 - Defining Academic Writing

English Development Team	Subject Specialists	Students
• An ability to synthesise numerous sources into a coherent piece of writing (paraphrasing) • A wide range of synonyms for use in academic work (paraphrasing) • A range of strategies for referring to authors inside the students' written work, to refer to research done (paraphrasing) • The ability to write reasonably coherently with regard to grammar • Paraphrasing/summary • An awareness of word forms and ability to manipulate them (paraphrasing) • Cohesion, linkers • AWL items • Conditionals • Complex noun phrases • Problem-solution-evaluation • Academic register • Compare contrast • Sentence syntax • Relative clauses • Understanding parts of speech (and their purpose) • Agreeing disagreeing opinion • Describing data accurately cf <u>trends</u> [original underlining] • Hedging language • Awareness of a variety of styles of academic writing • Proofreading/redrafting • Self-correction • Discourse structure • Paragraph structure • The writing process: planning / producing / reviewing	• Paraphrasing and referencing • Précis • Use of proofreading service • Sentence structure • Spelling • Grammar • Paragraphing • Punctuation • Essay and report writing (structuring) • Academic writing • Use of appropriate language • Academic vocabulary • Tenses • Constructing a persuasive argument • Writing skills (basic) • Writing in context	• Brain storming for establishing structure of essay/report • Mind diagram theoretical framework • Vocabulary – not only the words using in oral but also the vocabulary for writing academic assignment • Grammar • Synonyms and academic words

Despite this obvious discrepancy, in general, academic writing was considered important by all of the key stakeholders, most likely because writing tasks make up over half of the assessment tasks encountered by students at all levels, at 53.2% of all assessments used at MUC. The breakdown by level shows that at level 6, 60.6% of assessments are written, although this drops slightly to 51.3% for level 7. Despite this drop at level 7, the students at this level are expected to produce a 15,000-word

dissertation, which is worth 60 credits of their 180-credit programme, during their final 6 months in the UK, a significant undertaking in a foreign language. These figures reflect the literature, with many authors commenting on the importance of effective writing. Lillis and Scott (2007, p. 9) for example highlight that:

> students' written texts continue to constitute the main form of assessment and as such writing is a 'high stakes' activity in university education. If there are 'problems' with writing, they the student is likely to fail

AH in subject specialist group two summed up the general area of academic writing for his group, citing the following areas as being important for writing at HE level: "…things like sentence construction and phrasing, grammar, punctuation, tense, fluency, logic, vocabulary range [and] structure" (AH, Subject Specialist Group Two). These skills can be further broken down into the following subskills, based on subject specialist lecturer terminology: the basics, using the work of others, including paraphrasing, summarising and critical writing, and proofreading and redrafting of work.

6.4.1 The basics

Starting with 'the basics', this was an area which many of the subject specialists discussed at length. MW, for example, explained that for writing, students needed "basic writing skills, grammar, understanding sentence structure" (MW. Subject Specialist Group Two). In response to this, EC commented that:

> we have students come and they have no written skills. They basically if they write they're writing for text [SMS]
>
> EC (Subject Specialist Group Two)

This point was expanded by MW, who afforded it a great deal of importance. She emphasises that for students to be able to cope with their studies they need:

> [to be] able to write without text speak, but they can't even write an email that's clear. I would like them to be able to write in a way that I understand rather than [text speak]. I mean, that's relatively basic, isn't it?

<div align="right">MW (Subject Specialist Group Two)</div>

During the same discussion, AH explained that accurate "phrasing and use of language" (AH, Subject Specialist Group Two) were vital for students to be able to produce work which was of the standard he expected from HE level students.

This was an area also discussed by the first subject specialist group, with RB explaining that "I've got things about spelling...? Spelling, punctuation, grammar", adding later that in her experience for international students

> even just the in the real basic paragraphing and sentence, erm, capital letters and really basic stuff can be really weak at level four ... And level five and level six!.

<div align="right">RB (Subject Specialist Group One)</div>

In response to this CW commented that for him "sentence structure and grammar" were of particular importance.

To bring in the literature now, the comments above do not reflect the secondary findings, with the majority of authors explaining that subject specialists were more likely to focus on issues with organisation and discourse structure than the sentence level features. Bridgeman and Carlson (1984, p. 263) for example explain that

> in general, grammatical and sentence-level features (e.g., punctuation, spelling, sentence structure) were rated as less important than more global essay characteristics (e.g. paragraph organization, paper organization, quality of content, development of ideas).

This is supported by Evans and Morrison (2011) who highlight that issues with language were not considered as important by lecturers as was good organisation and

Donohue and Erling's (2012) comment referring to the fact that subject specialists tend to "look through" the language in order to find the meaning (see above).

The area of basic proficiency links to the need for greater vocabulary and grammar knowledge which was highlighted in particular by the students. This brings me again to the AWL, an area which has been returned to frequently throughout these findings. With regard to specific comments from the students, OM in the first student group highlighted that for her being able to use synonyms effectively was particularly important for her academic writing. In the second student focus group, CY mentioned that "vocabulary knowledge of writing is important" (CY, Student Group Two). In contrast, NX explained that for her it was grammar which was key; something which was also commented on by DY in the first student group. The importance which is afforded vocabulary and grammar knowledge by the students in the present study reflects the findings from many others. Student participants in a study reported by Andrade (2006) highlight limited vocabulary as being a particular issue for them, and in Evans' and Morrison's (2011) study, their student participants emphasise correct use of grammar, and an "apparent lack of a rich vocabulary" (p. 391) as being of particular importance. Their participants also believe that they have a "limited repertoire of sentence patterns" (*ibid*.).

As is evidenced in Table 19 (p. 218), vocabulary knowledge was listed as important by the English team, as would be expected, although their focus was more on specific academic vocabulary such as using hedging language, language relating to describing data and trends and specific vocabulary for introducing quotes and paraphrases. During their discussions, this area was returned to on several occasions. CL (English

Development Centre), for example, when discussing writing skills stated "I've got things like some grammar knowledge, knowledge of vocabulary".

What is clear, then, is that vocabulary and grammar knowledge are deemed important by all stakeholders when it comes to academic writing, and as with the participants in Lea and Street's (1998) study, the academic staff at MUC "have their own fairly well-defined views regarding what constitutes the elements of a *good* piece of student writing" (p. 162, my emphasis). There is, however, a difference here in that the subject specialists in this study are more concerned with grammar and vocabulary than those in other studies (see, for example, Donohue and Erling, 2012 and Evans and Morrison, 2011). Despite this difference, what the discussions are showing is that accurate use of language, clarity of expression and the ability to write according to genre-specific conventions (as was also seen in Chapter 5) were seen as a key starting point for students hoping to study at HE level in the UK.

In terms of defining the construct of writing, then, the first point for consideration is how well the assessments provide the students with the opportunity to demonstrate their proficiency in 'the basics' of vocabulary knowledge, grammar and discourse structure. In order to do this, it is necessary to investigate the generic grading criteria for the two programmes. The criteria award students marks for their lexical resource, grammatical range and accuracy and organisation (Appendix L), thus the broad grading categories clearly consider the basics. A closer look at the descriptors shows that without a good grasp of these basics, students will not be deemed to have the proficiency needed to be able to cope with their HE level studies at MUC. In fact, the 'fail' categories (D – F for the IEP and below 50% for the SEP) include comments relating to errors in word formation and spelling, illogical paragraphing and low levels of grammatical accuracy

which reflect the comments from the subject specialists as being issues with regards to students' academic writing. What is not clear, however, is what constitutes a 'basic' error, since the subject specialists' understanding and the English team understanding of this term could be very different; something which was evident from a comment made by RB (Subject Specialist Group One):

> But I'm really surprised meeting the students I meet in the final year, overseas students that have supposedly passed this, I'm amazed they've passed it. But, clearly they have passed it but then when I see them and when I read their written work, I can't believe they have passed it frankly. You know, so I wonder what your pass level's like. I know you've got grading criteria but it doesn't seem to tally with the standard of English or spoken English or written English…

This was reflected by SE (Subject Specialist Group Two) who commented:

> Yeah, because if they had just looking, thinking particularly from masters level, if they had passed all of this, we wouldn't have a lot of the problems that we have… You know, if they had passed it… Yeah. That if they were, you know, competent in all of these skills that are reflected in the assignment brief, then we wouldn't have a job. You know, we would do, but it would be a lot easier.

What can be inferred from this is that the English team and the subject specialists seem to have different expectations in terms of English language requirements, something which perhaps needs to be communicated better to all MUC staff. Certainly, there are issues with the grading criteria on closer scrutiny of aspects relating to the basics. For example, the difference between 'uses sufficient range of vocabulary' (75 – 84%) and 'uses an adequate range of vocabulary for this task' (60 – 74%) is unclear, since they are essentially synonymous terms. What is clear is that the requirements for a pass do not insist on perfect language use, since the criteria are based on a combination of IELTS criteria and the CEFR[11]. Since students need to demonstrate written proficiency equivalent to B2 on the CEFR or 6.0 on the IELTS scale, there is no requirement for

---

[11] The original criteria from which they are taken can be seen in appendices J and K.

them to produce consistently accurate writing; rather, what is required is that they can make their message clear despite inaccuracies in grammar, vocabulary, spelling and organisation.

Additionally, in terms of 'the basics' the descriptors for a pass on the HE programmes are equally subjective and potentially opaque as is seen in Table 20 below:

Table 20 - Descriptors for a Pass Grade on Assignments at MUC

Level	Descriptor
4	Basic attempt to present coherent work
5	Adequate standards of vocabulary
6	Adequate standards of vocabulary and grammar
7	Appropriate command of vocabulary and grammar leading to effective communication

What is clear, therefore, is that the subject specialist staff seem to have unrealistic expectations in terms of the written proficiency of the students, and what, on initial analysis, appeared to be potential issues with the reliability of the grading criteria used for the English assessments, seem instead to be due to these unrealistic expectations.

A further area for consideration with the IEP research report criteria is their ease of use. Although the key 'basics' are touched on in the descriptors, what could be questioned is how easily they can be applied. They are very complex criteria in that they offer two possibly competing sets of descriptors to be applied to one report, with weighted grades awarded. The first five areas of the criteria relate solely to the content of the report, that is the key sections required; the final three to the quality of language and academic writing skills. In considering the criteria in more detail in relation to 'the basics', the first issue is combining style and organisation in one area, with the introduction (and criteria relating to organisation) in a separate area. This makes the criteria difficult to apply in terms of the basics, and other areas as will be discussed below.

The rationale behind the complex criteria was to mirror the weighed approach taken to grading dissertations, and to give students experience of this approach in an attempt to enhance the authenticity of the grading experience. On reflection, however, the reliability of these criteria has been compromised by the attempt to enhance the authenticity, and review of these criteria is needed if they are to be truly reliable, and the corresponding assignments genuinely useful for all stakeholders.

6.4.2 Using the work of others

Building on the area of 'the basics' the next topic of discussion is using the work of others. As Cumming *et al.* (2000) explain "writing seldom calls for the writer to start completely from scratch but rather to draw on a variety of resources" (p. 3), and for the English team effective use of secondary sources was an important skill for students to show competence in. During these discussions it was often discussed along with note-taking and paraphrasing, as well as critical thinking, so the remainder of this section will discuss these areas together.

Starting with the English team discussions, BW commented that "paraphrasing and assimilating what they've learnt into their work" (BW, English Development Centre) are particularly important skills for coping with HE level study, and these are areas he returned to several times during the focus group, which reflects Weigle and Parker (2012) who explain that a key skill in academic writing is "discourse synthesis" (p. 119).

CL and LM also discussed paraphrasing, although they assigned it quite different purposes. CL explained that

> paraphrasing is … a way of using grammatical features and vocabulary features in a specific, for a specific purpose. And then there's the reading skills and the listening skills of note-taking erm which is more general for understanding purposes, whereas paraphrasing is more for er expressing yourself purposes, perhaps.
>
> CL (English Development Centre)

CL also commented "the ability to select information to communicate an idea which is a communicative skill rather than a purely more understanding based skill" (*ibid*.). In contrast, LM explained that for her, paraphrasing is more about students demonstrating understanding of what they have read than it is about their knowledge of vocabulary and grammar.

The ability to summarise and paraphrase the work of others was also something which was discussed by the subject specialist participants, with seven of the twelve participants mentioning it at some point during the focus groups. The exchanges below from the first group are typical of the discussions had at various points:

> AL: I've got something like paraphrasing and you know, differentiating theory to anecdotal theory which are more skills in academic writing.
> IJ: Paraphrasing, referencing and précis.
> CW: I've got reading and summarising.
>
> Extract from Subject Specialist Group One

Interestingly, although dealing with broadly the same area, in the first subject specialist group ST used slightly different terminology, explaining that for him what is important rather than effective paraphrase and summary is "being able to avoid plagiarism" (ST, Subject Specialist Group One). This implies that being able to use the work of others is more than just finding quotes and ideas and inserting them into the appropriate places in the work, and more than just changing one or two words in an attempt to paraphrase. What ST seemed to want from students is that they can use the work of others to support their own points rather than just repeating what is presented in the sources they consult. This is similar to comments from the literature, especially from Wiegle and

Parker (2012, pp. 119 – 120) who explain that, in relation to using the work of others in their writing

> students need to know when it is necessary to cite sources, when quotation is an appropriate strategy, and how to determine the boundaries between a close paraphrase and simply copying

Shi (2012) also touches on this area, commenting that in her study, "many students … believed that it was acceptable to copy long strings of words from a source text as long as the original author received credit" (p. 135). Thus, effective paraphrase is vital, and the key word here is 'effective', and what this actually means. This is, therefore, an area which must be considered when evaluating the writing grading criteria, as well as when considering the reading exams since this skill could be included in both sets of assessment.

In terms of the student participants, they all mentioned paraphrasing and summarising, with the ability to effectively manipulate language coming up in both groups. OM (Student Group One), for example, highlighted "how to do the quote and the synonyms" as well as "how to do the synergy" as being important skills she learnt on her English programme that are now useful to her current studies, thus emphasising how the English programme assessments aided with her "process of initiation … [into MUC's] academic discourse communit[y]" (Luna, 2002, p. 596).

To return to the literature, similar to the findings in the current study, Wiegle and Parker (2012) explain that "…the ability to synthesize ideas from sources into one's own ideas is an important component of academic writing ability" (p. 119), and Evans and Morrison (2011) highlight the importance of students being able to synthesise information from various sources. Jordan (2002) also emphasises the importance for

students not only to be able to summarise and paraphrase, but also to synthesise the ideas of the key authors in the field. Further, Wu and Hung (2011) highlight that a key requirement for students with regard to academic writing is that they "…employ their cognitive techniques in synthesizing and accumulating the information from [various] sources to write … independently" (p. 36).

What is arising from the discussions so far is that the construct of academic writing, needs to include paraphrasing and using the work of others to help support and develop ideas. That is not to say, as is evidenced in the literature review[12], that these issues are unique to international students, but since one learning outcome of these programmes is to equip students with the necessary academic skills to study up to postgraduate level in a UK higher education establishment, including these key academic skills are a compulsory component of both the programmes and the assessments used on them.

Coming to the assessments, paraphrasing is included in the reading and the writing assessments for both programmes, although how it was graded, especially in the reading assessments, might have questionable reliability. The use of secondary sources to support and develop a written argument is only included in the writing assignments.

Starting with the paraphrasing, the reading exams on both programmes include tasks which require students to use their own words to answer some of the questions. This is a direct test of their paraphrasing ability and thus these tasks have a high level of context validity. Although testing writing in a reading exam could impact on the construct validity of the assessment, the balance between context and construct needs to be considered carefully when establishing the overall predictive validity; what is clear is

---

[12] See, for example, Holder (1999), Lea (2004, 1998), Luna (2002)

that, on face value at least, these tasks have a high level of predictive validity since they are predicting how well students will be able to read, process and understand the information they are given and then put this information into their own words. Additionally, the intensive English programme exam includes a task which requires students to write a summary of the longest text in the exam. The instructions for the task include: "make sure you use your own words (paraphrase)" (IEP reading exam, 2010). Additionally, the mark scheme rewards students for using their own words and reformulating ideas. Since this is considered such an important skill by all of the stakeholders, it is clearly an important skill to be testing, although where best to test this skill could be questioned.

The writing assessments for both programmes overtly state that students must make use of secondary sources as is evidenced in the sections outlined below (Figures 35 and 36).

---

(4) **Findings and analysis.** This section evaluates the hotel chain's online resources, using the theory that you have identified in the literature. You must use paragraphs and signposting language to make your structure clear. You must also use references from the literature that you have read.

Figure 35 - Extract from SEP Writing Remit

Findings and analysis

- This must be divided into logical sections
- You must use headings to organise this section
- You need to use tables and charts to display your primary data clearly
- You should discuss what your findings indicate by using a combination of data from your questionnaire as well as books, journals and websites.

Figure 36 - Extract from IEP Semester 2 Writing Remit

Both remits explicitly require students to make use of secondary sources to support their discussion. This is further evidenced in looking at the grading criteria, with use of secondary sources being a separate category for both sets of criteria. On closer scrutiny,

however, the students are only rewarded for using secondary sources to support their points in the higher grades as is evidenced in the extracts below:

Table 21 - Extract from SEP Writing Grading Criteria

	Use of secondary sources
85 – 100%	Effective selection and use of support references and quotations conforming to college standards
75 – 84%	Mostly effective selection and use of support references conforming to college standards
60 – 74%	Good selection and use of references and quotations conforming closely to college standards

In terms of the defined construct, then, the writing assessments have a high level of validity. The majority of written assignments at MUC have some requirement for students to use the work of others to support their points, and this is particularly the case at levels 6 and 7, with students having to produce dissertations which require a large element of synthesis on the part of the students. The authenticity of the tasks, in that they closely reflect what is required on other courses at MUC, helps to enhance the context validity of the tasks and thus supports the predictions about students' abilities to synthesise and paraphrase the work of others in their writing. Additionally, in providing an authentic approach to the task, giving students the opportunity to develop their skills over time, the writing assessments could be deemed useful to all stakeholders in terms of using the work of others.

In addition to requiring the students to use the work of others in their reports, the two writing assignments also ask students to produce a literature matrix in which they need to synthesise the main ideas from their reading according to key themes that they deem useful for their assignment. This provides students with the opportunity to practice organising their reading into key themes prior to using it to support their writing. This

not only reflects the literature matrix which is an integral part of the dissertation at levels 6 and 7, but it is also a key skill for the students to develop.

6.4.3 Critical writing

Linked to paraphrasing, summarising and using the work of others is the idea of critical writing. This is a skill which is discussed at length in the literature, with Andrade (2010), Kingston and Forland (2008), Andrade (2006) and Kember (2000), amongst others all discussing it. Andrade (2006) in particular comments on the fact that, in terms of writing, tutors in a study she reported found that students "lacked critical thinking skills" (p. 138), adding that students are "unaccustomed to analysing the strengths and weaknesses of an argument" (*op. cit.*, p. 139). To come now to the primary findings, during the first subject specialist group, the categorisation activity raised the idea of the ability to analyse the written word, with four of the participants including it on a sticky note and TM (Subject Specialist Group One) commenting at the end of the activity "…we'd all perhaps agree critical thinking is important". More specifically, in this was an area emphasised by ST when he explained that:

> I think our expectation is that they should be able to come in and critically analyse, [but] I think our experience shows that's not necessarily the case.

> ST (Subject Specialist Group One)

TM (Subject Specialist Group One) also discussed the fact that she expects students "to take a questioning approach" when writing, with CW (Subject Specialist Group One) adding "I think study skills is about asking questions". This reflects the findings from Kiely (2004) who explains that in higher education the term critical is frequently used in assessment criteria and feedback, although he found that its meaning is rarely explained to students.

In the second subject specialist group, the same activity led to participants identifying drawing logical conclusions and analysis and evaluation being highlighted as key. MW, in fact, found the following skills of particular importance, yet skills which students lack: "your analysis and evaluation, your critical reading and being able to draw conclusions" (MW, Subject Specialist Group Two), with KB (Subject Specialist Group Two) adding when discussing masters students in particular "...you might be looking for more critical discussion".

As with effective paraphrasing, the issues highlighted in relation to critical skills are not unique to international students. However, the learning outcomes for the programmes encourage students to develop knowledge and understanding of approaches to study / culture in a UK higher education establishment. Thus, since the critical elements are such an important component in Western education (Lea, 1998), and are a key element of the programme learning outcomes, it is important that the EAP assessments being investigated here include elements of training in these areas in order to fully prepare students for their future studies.

The English team also discussed the importance of critical writing, with SD in fact explaining that:

> when I first came here, I think I would've thought along those lines of writing, speaking, listening, but the longer I've been engaged in research myself, and supervising dissertations and marking dissertations I've sort of drifted more towards thinking, well, the critical component is a very important aspect, and usually lacking.

SD (English Development Centre)

LM also touched on the idea of criticality in writing, explaining that criticality is an area with which students need help, particularly in their writing, adding that the assessments are a good place in which to help them develop this skill.

This area was only briefly discussed by the student participants, although OM (Student Group One) did discuss "how to do the analysing … how to analyse the literature" as a key writing skill, and DY commenting in response to this that using sources critically was a new skill for her. In considering the assessments, it will be the writing assignments where the students will be most likely to demonstrate this critical thought, and as such I will consider how the writing grading criteria assess students' ability to do this, if at all.

Looking at the writing assignments, students are provided with opportunities on both courses to demonstrate their analytical and critical writing skills. Starting with the IEP, the second semester writing task in particular includes the following instructions (Figure 37 on the following page) for students about what to include in each section of their report. These detailed instructions clearly request that students take a questioning approach, and include analysis in their writing, two key areas of academic writing expected by the subject specialists. By requiring students to explain why they made certain choices, and by asking them to comment on their primary findings in relation to the secondary data, they are being provided with the opportunity to develop, and then demonstrate, key critical writing skills.

Introduction

This needs to include:
- A clear statement of aim explaining what the report is doing
- A clear rationale explaining why the report is being written
- The points that will be covered
- The order the points will be covered

Methodology

This needs to state
- Where the secondary data came from (books, internet, for example)
- How you did the primary research
- How many questions you asked and why
- Which questions you asked and why
- How many people you asked and why

In this section, you must also use information from your research methods reading pack to defend and justify your choices.

Findings and analysis
- This must be divided into logical sections
- You must use headings to organise this section
- You need to use tables and charts to display your primary data clearly
- You should discuss what your findings indicate by using a combination of data from your questionnaires as well as books, journals and websites.

Conclusions and recommendations
- A summary of your key findings
- Sensible recommendations for students and institutions based on the conclusions

Appendices

This section includes extra information. For this report you **MUST** include:
- A thematic literature review matrix which synthesises the sources you are using
- A literature evaluation matrix which critically reviews some of your key sources
- A copy of your questionnaire

Figure 37 - Extract from IEP Semester 2 Writing Remit

Considering the grading criteria which relate to this particular assignment (Table 22), the following statements provide the raters with the chance to reward students for their critical and analytical writing. These descriptors effectively operationalize the concept of analytical writing.

Table 22 - Extract from IEP Report Grading Criteria

A	B	C	D	E
Good evaluation of secondary sources	Adequate evaluation of secondary sources	Some evaluation of secondary sources, but this may be quite simplistic in nature	Only a basic evaluation of some of the sources / not all sources have been evaluated	Secondary sources have not been evaluated
Effective justification of approaches based on research methods texts	Good justification of approaches based on research methods texts	Some justification of approaches based on research methods texts	Limited justification of approaches	Approaches have not been justified
Comparative discussion of primary findings with secondary	A good attempt to compare primary and secondary findings but this is not always successful	Some comparison of primary and secondary findings, but this is limited	Limited comparison of primary and secondary findings	No comparison of primary and secondary findings

The SEP also provides the students with the opportunity to develop and demonstrate their ability to write critically and analytically, as is seen in the writing assignment extract on the following page in Figure 38. As with the IEP writing assignment, students are provided with the opportunity to develop and subsequently demonstrate their analytical writing ability. Unlike in the IEP assignment, however, students are not rewarded for this in the grading criteria; an area which needs addressing.

Overall, in terms of the critical and analytical writing skills, the students on both programmes are given the chance to develop and demonstrate their ability to do this. What is lacking is consistency in terms of grading criteria descriptors, with the SEP descriptors requiring amendment to address this.

Online resources (for example, websites) present hotels with new opportunities to market themselves to customers. Use the AIDA model to evaluate the online resources of **one** of the following hotel chains:

(1) Accor Hotels
(2) Premier Inn
(3) Hilton Hotels
(4) Novotel Hotels

You need to evaluate the resources in relation to the three following scenarios:

a) A family looking for a weekend break
b) A business looking to hold a conference
c) A couple looking to visit the restaurant for a meal

Your report should have these sections:

(1) A brief **abstract**, summarising the main sections of the report and presenting the main findings and recommendations. The abstract is not included in the word count.
(2) A clear **introduction**, outlining the aim of the report and its structure.
(3) A **literature review**, outlining the AIDA theory and its strengths/weaknesses. Note that you do not need to include all the literature from your literature matrix – only the useful literature.
(4) **Findings and analysis**. This section evaluates the hotel chain's online resources, using the theory you have identified in the literature. You must use paragraphs and signposting language to make your structure clear. You with must also use references from the literature that you have read.
(5) A clear **conclusion**, where you summarise the key points from the analysis section, and present **recommendations** for the hotel to modify its current practices.
(6) An **appendix** with the literature matrix arranging the information from the articles and textbooks you have read into themes and giving a brief evaluation of the different sources. The appendix is not included in the word count.

You can select appropriate diagrams and pictures to use in your literature review and analysis section if appropriate. Make sure that you provide clear references for any pictures and diagrams, though.

IMPORTANT!
Please note – you can **only** reference materials included in the reading pack. You do **not** have to read every article from beginning to end. Instead, you should read **relevant** sections of the articles in order to help you write your report. You will need to identify what is relevant and what isn't.

Figure 38 - Extract from SEP Writing Remit

Coming the SEP grading criteria (Table 23 on the following page), these offer the following comments relating to criticality, with particular reference to the literature matrix:

Table 23 - Extract from SEP Report Grading Criteria

Grade	Use of secondary sources
85 -100%	Effective selection and use of support references and quotations conforming to college standards A clearly organised and well-synthesised literature matrix Good evaluation of secondary sources
75 – 84%	Mostly effective selection and use of support references and quotations conforming to college standards A clearly organised and well-synthesised literature matrix Mostly good evaluation of secondary sources
60 – 74%	Good selection and use of references and quotations conforming closely to college standards An organised matrix with some evidence of synthesis Adequate evaluation of secondary sources
50 – 59% (Borderline fail)	There is a genuine attempt to reference according to college standards A relatively well organised matrix with some evidence of synthesis Some evaluation of secondary sources
40 – 49% (fail)	There is a genuine attempt to reference according to college standards An attempt made to organise and synthesis sources in the matrix Some evaluation of secondary sources
35 – 39% (fail)	Limited attempt to comply with the college standard Limited synthesis of sources in the matrix Only a basic evaluation of some of the sources/not all sources have been evaluated
Below 35%	Secondary sources have not been evaluated No synthesis of sources

As can be seen, it is only in the 'use of secondary sources' area that any element of criticality is considered, and this is only in relation to the literature matrix. In terms of representing the construct, then, these criteria are lacking and the students are not rewarded for their critical or analytical writing skills at all. Further, although the criteria are reliable in that the same areas are covered in each grade band, the lack of validity in terms of the comments relating to analytical writing impact significantly on the overall usefulness of this writing assignment.

6.4.4 Summary comments

Overall, the writing assignments appear to offer students with sufficient opportunity to demonstrate their competence in academic writing in a suitably academic context, hence on face value they have an adequate level of context validity. Despite the comments from the subject specialists regarding the presentation of the assignments, their

presentation closely mirrors that of other assignments used at MUC, thus they also have high face validity. The construct, that is writing, appears to be adequately assessed, although the main points of contention come from the reliability of the grading criteria, which either do not offer consistent grade descriptors across the different bands, or do not suitably operationalize the wider academic skills.

6.5 What is academic speaking?

Along with writing, speaking skills are highlighted by Andrade (2006) as one of the sets of skills which have the "great[est] effect on achievement and adjustment" (p. 148) in higher education. By way of definition, Butler *et al.* (2000) explain that speaking in general is "the use of oral language to interact directly and immediately with others" (p. 2), defining *academic* speaking as "such interactions as occur in academic settings" (p. 2). These academic settings involve both in-class encounters as well as encounters with others outside the classroom, including both formal, assessed tasks, as well as informal speaking tasks. The focus group participants identified the following speaking skills during the sticky note activity (Table 24 on the following page). Interestingly, this table shows that the student participants placed slightly less importance on the speaking skills than did the teachers, both English and subject specialist. In terms of the construct, then, the following broad areas arose from the focus groups: general communication skills, participating in class, working as part of a group and academic speaking tasks which include seminars, presentations and debates.

Table 24 - Defining Academic Speaking

English Development Team	Subject Specialists	Students
- AWL items - Conditionals - Knowledge of subject-specific technical vocabulary - Knowledge of academic vocabulary – awareness of some of the subtleties - Agreeing disagreeing opinion - Presentation skills - Clarification/comprehension - Referring to research in a debate - Turn taking - Fluency - Self-correction/proofreading - Discourse structure - Structuring presentations - Understanding different genres of written/spoken texts	- Oral communication - Debating - Questioning approach - Presentation skills - Grammar - Sentence structure - Use of appropriate language - Academic vocabulary - Working with others - Expressing views verbally - Group work/interaction - Confidence in speaking to share opinions - The art of debate	- Present a ppt - Oral expression (questioning and communicating with course lecturer) - Speaking - Asking questions

The majority of these areas have already been discussed in some detail in Chapter 5. In particular the areas of the construct which relate to participating in class, working in groups and carrying out academic tasks such as seminars and presentations. With this in mind, this section will just consider just fluency.

6.5.1 Fluency

Ferris and Tagg (1996b, p. 311) in their study find that

> in general … with regard to speaking skills, professors were most concerned with ESL students' ability to express themselves fluently in class discussions and to give clear, coherent answers to questions asked of them.

Considering this same point, RB commented that students:

> don't seem to have the really basic fluency through not having the opportunity to speak to lots, to English people.
>
> RB (Subject Specialist Group One)

This was picked up by AL who commented that:

> I've got oral communication because it's ok to write, some people can write it down by they can't actually articulate what they're doing.
>
> AL (Subject Specialist Group One)

These ideas were reflected in the second subject specialist focus group too, and MW, for example, commented "you know, we're dealing with people who can barely, you know, talk to you" (MW, Subject Specialist Group Two). SE added to this, highlighting that in her experience, many international students struggle to "string a sentence together" (SE, Subject Specialist Group Two). What this shows is that for the subject specialists at least, general fluency is particularly important when it comes to oral communication, and this implies that fluency is, perhaps, more important than accuracy. The English team also discuss fluency, albeit briefly, as is seen in the following exchange:

> CL: But fluency is about using grammar effectively in speech, isn't it?
> LM: You can use grammar effectively in speech and not be fluent
> SD: [Makiko] springs to mind here.
>   Extract from English Development Centre

This is an interesting exchange, which needs some further explanation. The example student given by SD was a student who we all taught and who spoke with near perfect grammatical accuracy. This grammatical accuracy, however, severely impeded her fluency since she spent so long formulating sentences or responses to comments and questions that it was difficult to follow her points. This student is a good example of the points being made by both the subject specialists and English team participants. In taking so long to formulate sentences, this gives the impression that she lacks the 'basic fluency' deemed so important by RB; this lack of fluency would give the impression that she had trouble 'stringing a sentence together'. Interestingly, the student participants did not comment on the areas of fluency or basic accuracy in terms of the more informal speaking tasks.

In terms of the literature, these 'basic' communication skills are also discussed by a number of authors. Hyland and Hamp-Lyons (2002), for example, emphasise that

EAP should include "not only study skills teaching but also a great deal of what might be seen as general English as well" (p. 2), and Dooey (2010) highlights that general English skills and informal oral communication skills are key. In terms of fluency in particular, Evans and Morrison (2011) highlight communicating ideas fluently as being important.

In terms of the assessments, there is no direct mention of fluency on the assignment remits at all, however, it is included in the grading criteria used for all speaking assignments:

Table 25 - Speaking Grading Criteria Operationalization of Fluency

Grade	Fluency	Use of notes
85 -100%	Speaks fluently with only rare repetition or self-correction any hesitation is content-related rather than to find words or grammar.	Delivered with no reference to notes
75 – 84%	Speaks at length without loss of coherence; may demonstrate language-related hesitation at times	Delivered with few references to notes
60 – 74%	Is willing to speak at length, but may lose coherence at times due to occasional self-correction or hesitation	Occasional references to notes
50 – 59% (Borderline fail)	*No descriptor relating to fluency*	Delivered with reference to notes
40 – 49% (fail)	Usually maintains flow of speech., but uses repetition, self-correction / slow speech to keep going	Delivered with reference to notes
35 – 39% (fail)	Cannot speak without noticeable pauses and may speak slowly with frequent repetition and self-correction	*No descriptor relating to use of notes*
Below 35%	*No descriptor provided*	Contribution appears scripted

Although the speaking grading criteria do go some way to operationalizing fluency, this is inconsistent, hence the reliability of the criteria is instantly called into question. For these criteria to be at all reliable in terms of the speaking construct, uniform use of terminology across all levels is required, and fluency needs to be included in all grade bands. A further issue is identifying what is 'language-related hesitation' as opposed to hesitation that is 'content-related'. In order for these criteria to be useful to raters,

examples of the different levels could be included (see recommendations in Chapter 7 below).

Coming to the comments relating to use of notes, the first possible issue is that highest level descriptor "delivered with no reference to notes". It is not clear why this approach should be deemed worthy of a high level score, since students are expected to make reference to previous research during their discussions. I would recommend that the criteria be amended to 'notes used effectively to support discussion'. In doing this, it is clear that students are expected to use notes, and they are rewarded for doing so effectively, rather than being unfairly rewarded for not preparing sufficiently for the discussion.

6.5.2 Summary comments

The construct of academic speaking as outlined both here and in the previous chapter is quite complex. Essentially, in an academic context, students are expected to be able to give formal presentations, participate in seminars and debates, ask and answer questions in class and work with group members to produce coursework. In order to do this, they need to be clear, fluent speakers who have sufficient vocabulary to be able to express themselves effectively in these situations. As discussed in Chapter 5, the speaking assessments are quite limited in that they only provide students the opportunity to practise seminar style discussions, hence in terms of construct and context validity they are limited since they only address one area of academic speaking. The rationale for this, however, comes from the need for students to be able to speak fluently, and spontaneously, thus, in this area they could be considered valid since they require students to do this. Assessing this fluency in presentations is more difficult, since this type of speaking task is easier to script. Thus using only seminars is an appropriate

choice in terms of assessing how quickly students can formulate and express their ideas without a script. In this respect, then, the speaking assessments are valid. The reliability of the grading criteria in respect to fluency, however, is questionable since the criteria lack consistency, and in places are missing key information. Thus, in applying the usefulness framework set out in Chapter 3, it is clear that the speaking assessments require significant development to ensure that they are truly representative of the construct of academic speaking and to ensure the grading criteria are reliable.

6.6 Summary

Throughout this chapter, I have applied the evaluative framework developed for assessing the usefulness of in-use in-house EAP assessments, concentrating on the construct, context, content and, where appropriate, face validity, as well as considering the reliability of the grading criteria and mark schemes. Based on this evaluation, it is very clear that there are a number of issues with the assessments in use in terms of their usefulness.

Of particular note is the lack of consistency in terms of assessment across the different programmes. Further, the predictive validity of the battery of assessments is questionable, with each assessment having a series of strengths and weaknesses. The main area of weakness with all assessments, however, has been noted as the reliability, with mark schemes and grading criteria demonstrating several areas of weakness.

In completing this final discussion chapter, I have now fully satisfied objectives six and seven of this study. Objective six has been achieved with the comments from the stakeholders being used to extend the framework and provide the key details relating to what is an authentic experience as well as what constitutes the construct and context. In

expanding and applying the evaluative framework, I have been able to establish the

overall usefulness of the EAP assessments in use at MUC between 2010 and 2014.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

7.1 Review of the research question and the main findings of the study

This study set out to investigate how EAP practitioners might go about establishing the usefulness of in-use, in-house EAP assessments. In order to do this, based on the literature review, an evaluative framework which considers authenticity, validity and reliability was developed and then demonstrated. The key findings from this study have been summarised at the end of each chapter, and are now reviewed below.

- EAP is more than just language; in helping students adapt to academic life in not only a new institution, but also a new country, EAP *must* also include elements of both academic and institutional acculturation. This much is clear from the review of literature in these areas which considers approaches to teaching, assessment and classroom behaviour (Cheng and Erben, 2012; Dooey, 2010; Smith and Khawaja, 2010; Wong and Thang, 2008; Andrade, 2006; Watson Todd, 2003; Jordan, 1997), as well as the potentially huge difference in terms of the overall teaching philosophy of the country (Ramachandran, 2011; Kingston and Forland, 2008; Kember, 2000). What this thesis argues, however, is that these also need to be reflected in the assessments used on pre-sessional language programmes. If the assessments, as pre-sessional assessments often are, need to be used to predict how well students will cope in their new academic environment, then they must provide students with the opportunities to both develop and demonstrate their competences in these areas. In order for in-house developed assessments to do this, EAP practitioners need to work in conjunction with the subject specialist tutors, also considering the experiences of students, to

carry out regular audits of the assessments so as to ensure their continued usefulness. This study has provided not only a framework for this purpose, but has demonstrated how it could be applied.

- The language which students produce as a result of the assessments they do needs to reflect the language that they will produce when they progress onto their main programmes of study. The authenticity of assessment tasks as discussed in the literature review is not a new idea, with several authors highlighting the importance of linking the assessment tasks to 'real world' situations, that is, the situations which reflect academic life as they will encounter it in their new insitution (Malone and Montee, 2014; Donohue and Erling, 2012; Douglas, 2010; Watson Todd, 2003; McNamara, 2000; Bachman and Palmer, 1996). The real-world situations to which in-house developed EAP assessments are related are much easier to establish than those to which large-scale examinations are related, and, again, this thesis has provided a framework for establishing these. What has emerged throughout this study, particularly when discussing the assessments with the students and subject specialists, is that it is vital that the language produced for and tested through EAP assessments mirror the real world language use tasks. If it does, then the assessments can be deemed useful assessments of academic language.

- Validity is an area which is widely discussed in the literature, with several authors preferring construct validity as an overarching term (Lissitz and Samuelson, 2007; Bachman and Palmer, 1996; Messick, 1989). As discussed throughout this study, however, I have taken a different approach, preferring predictive validity. Since one of the purposes of the assessments being investigated is to make predictions about how well students will cope in their new environment and with the various tasks they will face, predictive validity

seemed a more helpful term for establishing assessment usefulness. In terms of the evaluative framework, the predictive validity needs to be established by considering various types of validity evidence as recommended by Al Hajr (2014), Moss (2007), Sireci (2007) and Weir (2005). In looking at the construct, content, context and face validity of the assessments, and in discussing these points with the various stakeholders, the usefulness of the assessments in use was further established.

- The reliability of assessments is something which is often established using complex statistics and various versions of tests, and it is more usually carried out during the assessment development stage (Fulcher, 2010; Weir and Wu, 2006; Weir, 2005; Hughes, 2003; Bachman, 1990; Heaton, 1975). Since the purpose of this study was to investigate in-use tests, clearly this approach was not appropriate. Further, since I aimed to provide an evaluative framework which could be used by EAP practitioners, then an approach to establishing reliability which did not require large amounts of statistical analysis was likely to be a more suitable approach. In considering instead the reliability of the mark schemes and grading criteria, this study has demonstrated how an EAP practitioner might be able to establish the reliability, and therefore the overall usefulness of in-house developed assessments.

To review the research question and objectives more explicitly, this study set out to establish how EAP practitioners could establish the usefulness of in-use in-house EAP assessments. To achieve this, seven objectives were set, and I will now outline how each of these has been achieved.

- **To provide a brief history of English for Academic Purposes**

In the literature review in Chapter 2, I established that EAP as a specific discipline began in the 1970s, emerging from the specific academic language needs arising from the increasing numbers of international students studying in English-speaking institutions. In outlining the development of the discipline since the 1970s, the literature review showed that teaching EAP has developed in terms of teaching practices, teaching materials, the use of technology, assessment methods and teacher training. By way of a summary, EAP was defined for the purposes of this thesis as the creation or interpretation of intended meanings in *academic* discourses by an individual, or as the dynamic and interactive negotiation of intended meanings between two or more individuals in various *academic* situations.

- **To critically evaluate the literature relating to language assessments, assessment evaluation and academic acculturation**

The discussions in Chapter 2 regarding assessment, and more specifically assessment in HE show that assessment can, and indeed should, serve many purposes including, but not limited to, evaluating knowledge, judging mastery of skills, establishing what has been learnt of a syllabus, as well as encouraging further development through formative feedback (Lau, 2016; Pereira *et al.*, 2016; Fletcher *et al.*, 2012; Price *et al.*, 2011; Dunn *et al.*, 2004; Walkin, 1990).

The literature also indicates that assessment in HE is in a state of flux (Pereira *et al.*, 2016, Knight, 2002; Black and Wiliam,1998; Hager and Butler, 1996), especially in terms of types of assessments being used as a result of pressures from students, institutional management, industry, policy-makers and increasingly stretched resources

(Hefce, 2016; Lau, 2016; Clements and Cord, 2013; Price *et al.,* 2011). Despite these assertions, findings from the primary research indicate that, at MUC at least, the written assignment still dominates, along with presentations and seminars.

Chapter 2 discussions also consider language assessments, defining them at their most basic as being an instrument with which to measure language ability and evaluate language performance (Heaton, 1975; Douglas, 2010). Further discussions, however, demonstrated that language tests can be very complex owing to their multifaceted nature (Davies, 1990), which sees them being formative and summative at the same time, whilst also being used as predictors of future performance (Douglas, 2010; McNamara, 2000; Davies, 1990) and hence key decision-making instruments in terms of progression (Davies, 1990). Additional complexities are caused by the fact that language assessments are indirect measures of language ability (Douglas, 2010) and almost impossible to grade with exact numerical accuracy (Douglas, 2010; Davies, 1990).

Certainly, in investigating the assessments used as the focus of this study, the complexities of EAP have become apparent, with the multifaceted nature of the assessments in terms of being formative *and* summative, as well as needing to aid academic acculturation becoming clear as the discussion progressed. Additionally, in investigating the mark schemes and grading criteria, the need to be precise in terms of operationalising skills and abilities clearly, and reducing the potential impact of subjective grading emerged as key since it is not possible to grade proficiency on a precise numerical scale.

The academic acculturation element is considered in both Chapters 2 and 3. However, the discussions of academic literacy have demonstrated that academic acculturation is something which is necessary for all students (Goldingay *et al.*, 2016; Lea, 2004), and that viewing international students as a homogenous group, all of whom are in deficit when compared to home students, is not necessarily helpful or even accurate (Hammersley-Fletcher and Hanley, 2016; Durkin, 2008; Ha, 2006; Kumaravadivelu, 2003). Despite this, academic acculturation is seen to be a key element of EAP in terms of preparing students for life in their new institution, and is thus a key factor when considering the usefulness of EAP assessments (Bang and Montgomery, 2013; Ramachandran, 2011; Smith and Khawaja, 2011; Kingston and Forland, 2008; Kember, 2000). In terms of the findings of this study, there was general agreement that the academic acculturation element is key in EAP, as well as in EAP assessments, with all stakeholder groups including acculturation elements in their discussions. However, despite the literature indicating that viewing international students in deficit is something which should be done with caution, the findings from the key stakeholders indicate an institutional culture which does just that.

- **To develop and evaluative framework to establish the usefulness of EAP assessments**

Heavily influenced by the works of Bachman (1990) and Bachman and Palmer (1996), Chapter 3 discusses various approaches to establishing how usefulness of language assessments, and culminates in an evaluative framework designed specifically for use with in-use, in-house developed EAP assessments. The framework incorporates three areas for analysis: authenticity, predictive validity, and reliability, thus not wholly conforming to the test usefulness framework outlined by Bachman and Palmer (1996).

The authenticity is concerned with how well tasks included in the assessments being evaluated reflect the reality of academic life at a chosen institution; predictive validity is established using various forms of evidence which consider the construct, content, context and face validity of the assessments. Reliability in this framework concerns not numerical or statistical measures, rather it is established by examining grading criteria and mark schemes.

- **To investigate EAP staff, student and subject specialist staff opinions regarding the language and wider skills required for coping with HE level study**

By adopting a case-study approach, this thesis investigated the opinions of three key stakeholder groups using five focus groups as outlined in the methodology in Chapter 4. As well as discussions, free-listing and pile-sorting activities were used with all focus groups as a way of establishing institutional requirements and expectations of students studying at HE level, as well as establishing key institutional PATs and identifying typical TLU situations.

- **To carry out an audit of subject-specific assessments in use at the case-study HEI to establish typical assessment tasks in use between 2010 and 2014**

An audit of the assessments used on HE programmes at MUC demonstrated that the most frequently used assessment tasks remained essays and reports, with more practical assessments continuing to play a more limited role, despite the literature review indicating that HE assessment is in a state of flux. In relation to the levels of study that EAP students most frequently progress onto, the audit demonstrated that reports were

more frequently used as a method of assessment than were essays, but essay-writing remained an important skill.

- **To use the opinions of EAP lecturers, subject specialist lecturers and former EAP students along with the results of the assessment audit to define and expand the key terms in the evaluative framework**

The discussions in Chapters 5 and 6 are based around the evaluative framework presented in Chapter 3. The comments and the results of the focus group activities have been used to establish what constitutes authentic listening, reading, writing and speaking situations. They have further been used to define the constructs of academic listening, reading, writing and speaking. Stakeholder comments were also used so as to establish the face validity of assignments, as well as to establish the specific institutional context. In using the primary data in this way, it was possible to further refine the evaluative framework for application in the specific context of the case-study institution.

- **By amalgamating the findings from the primary and secondary research, to critically evaluate the EAP assessments being used at the case study HEI between 2010 and 2014**

The framework as expanded with the use of the focus group findings was applied in Chapters 5 and 6, which presented a critical evaluation of the EAP assessments in use at MUC between 2010 and 2014. The evaluation of the assessments has identified both strengths and weaknesses in the assessments, with the weaknesses calling into question the overall usefulness of the assessments. Of particular concern was the reliability of mark schemes and grading criteria. In general, the critical evaluation indicated that a number of improvements were required with all of the assessments.

In achieving these objectives, the study's research question has been addressed, and this study has essentially developed and subsequently applied an evaluative framework to a suite of in-use, in-house EAP assessments. The overall conclusion of the application of the framework is that, despite the strengths of the writing and speaking assignments in particular, the assessments overall cannot be deemed useful since the weaknesses of different tasks, and the limited reliability of mark schemes and grading criteria call into question any grades awarded on the basis of the assessments used. As a result, significant development in all of the assessments is required.

7.2 Concluding comments

Considering the above summary comments and review of the objectives, it is clear that getting EAP assessment right is vital for several reasons. The assessments act as key determiners to academic success. If the students do not achieve on these assessments, then their academic careers are either over, or at least delayed, so the assessments need to be reliable instruments for establishing the students' language proficiency. Moreover, the assessments need to be testing the right *type* of language, in the right contexts if the predictions made on the basis of student performances on them are to be valid. In addition to their validity and reliability, however, language tests, as was discussed in Chapters 2 and 3 have a great deal of impact on curriculum development and design as well as student academic acculturation, areas which need to be considered when deciding on their usefulness, and the potential need for further assessment development.

EAP is complex to define owing to its multipurpose nature; a complexity which follows through into the assessments used to test it. As a field of research, EAP goes back to

the mid-1970s, when the term was first coined (Ypsilandis and Kandaridou, 2007; Benesch, 2001). To return to Hyland and Hamp-Lyons' (2003, p. 3) definition, EAP:

> seeks to provide insights into the structures and meanings of academic texts, into the demands placed by academic contexts on communicative behaviours, and into the pedagogical practices by which these behaviours can be developed.

The findings from the literature review indicate that students view their English programmes as training grounds for their future studies (Dooey, 2010), thus the assessments used on them need to offer further training for students in terms of their academic acculturation. Additionally, the subject specialist lecturers appear to view the English language programmes at MUC as a 'fix all' which will cure students of all their academic ills before they commence their studies. Further, the more recent changes in HE assessment practices reported by Boud and Falchikov (2007), Norton (2007) and Clegg and Bryan (2006) mean that courses which prepare students for access to HE level study need to remain current. Subsequently, the assessments used on such courses need to reflect these changes.

The focus group discussions with the three key stakeholder groups revealed a number of areas which are important for inclusion in EAP assessments in terms of how the assessments can help students with academic acculturation, as well as how best to assess the academic language deemed appropriate for HE level study. The investigation of the assessments on the basis of the evaluative framework presented at the end of Chapter 3 in conjunction with the focus group discussions highlighted both the strengths and weaknesses of the assessments in use at MUC between 2010 and 2014, providing suggestions for improvement at the end of the discussion chapters, which I will now follow.

7.3 Recommendations

The discussion of the assessments in use at MUC between 2010 and 2014 has revealed several areas of weakness, highlighting the need for significant changes. The purpose of the evaluative framework, as outlined in Chapter 3, is to establish whether changes need to be made to assessments to enhance their usefulness. In applying this framework, not only have I demonstrated how this can be approached, I have also highlighted where the assessments need to be improved and how. This brings me to my recommendations, which explain how MUC can make improvements to the assessments:

- Since academic acculturation is a key aim of EAP programmes, as well as an expectation from subject specialist staff and students, the assessments need to reflect the reality of study at MUC. Tasks should be refined, or redesigned, to ensure that they match those which students will encounter on progression, so as to enable the EAP staff to make valid predictions about whether students will be able to cope with them.

- MUC (and any other institution applying this framework) needs to outline a set of key academic competencies which can then be integrated into the EAP assessments, providing the students with the opportunity to develop these.

- The assessments were shown to be lacking in the representation of the AWL, implying limited content validity in terms of vocabulary. It is recommended that listening and reading texts in particular be amended to increase the coverage of the AWL, and reduce the amount of off-list language included.

- In terms of predictive validity, these assessments vary in their quality. Each has strengths and weaknesses; thus the most useful tasks from each assessment should be developed, with those tasks which have been deemed to have issues, in terms of content, context, construct or face validity, being replaced with tasks

which more closely resemble those which students will have to do when they progress. The speaking and writing assignments seem to be the most useful, owing to the high levels of face validity, construct validity and content validity. Essentially, these assignments most closely represent what students will be expected to do when they progress. Changes in areas such as including PowerPoint slides in the listening assessments would enhance the face validity, and ultimately aid students with their academic acculturation, thus enhancing the usefulness of the assessments in that regard.

- With all the EAP programmes having the same learning outcomes, it should follow that similar assessments be used on all of them. Thus, since the assessments are supposed to be assessing the same skills, for the same purpose, at the same level, the exams and assignments used on these programmes should be designed so that they include the same sets of tasks.

- A final task, or tasks, linked to the listening note-taking task(s) is an authentic way to further assess comprehension, and I recommend that a task which requires students to write a summary of the lecture is included in all listening assessments since this reflects the reality of students using ideas from lectures to support their writing.

- The writing assessments provide students with ample opportunity to develop as more autonomous learners, and as such this element should remain unchanged, although reading packs could be shortened, or students provided with reading lists only to give them the opportunity to develop online and library research skills.

- The opportunity for students to develop discussion skills gradually through a series of small, recorded tasks on the IEP helps students gain in confidence in

sharing their opinions. I recommend that this be included in SEP assessments too.

- Although testing speaking through seminars alone provides the students with the opportunity to practise an approach to oral assessment which might be unfamiliar, and which provides English staff with a truer representation of their spontaneous speaking ability, MUC tends to make greater use of presentations. I therefore recommend that the speaking assessments include a presentation task as well as the seminars.

- The biggest area of weakness for all assessments was the reliability of mark schemes and grading criteria. Subjective marking played too big a part in the grading process on all assessments, with insufficient examples and descriptors provided for raters. In order to improve this particular area, I recommend that raters be provided with consistent speaking and writing mark schemes, which effectively operationalize all aspects of language use and academic acculturation across all grade bands. Additionally, exam mark schemes need to include more explicit examples and breakdowns of how marks should be awarded. In doing this, the University can be more certain that the marks awarded on the programmes are more reliable.

Taking the above comments into consideration, by applying the evaluative framework from Chapter 3, it has become clear that action research in the form of amending and checking the assessments is required.

7.4 Limitations and recommendations for further study

Although an in-depth study of the practices at one institution, this thesis does still have some shortcomings. Indeed, I managed to get the opinions of just four students, which

might not be considered a truly representative sample of the student population. Future studies, therefore, could seek to involve more students in the research. This could be done through the use of a questionnaire rather than attempting to encourage more students to participate in focus groups since this approach has not been as fruitful as it could have been. Another alternative could have been to interview the students individually at a later date rather than ask them to participate in focus groups, however, as was discussed in Chapter 3, this could have had ethical implications in terms of treating the students differently to the other participants.

This thesis, as explained above, has contributed an evaluative framework which can be applied in different HEIs, yet here I consider a smaller, specialist institution. Applying this framework at a larger institution would be a further test of how well it works in practice, and indeed a comparative study applying the framework across various institutions could lead to results which form the basis of wider changes in the field of EAP assessment.

In terms of the model itself, I have considered reliability from a mostly qualitative point of view. Including a statistical test of reliability within the model, as discussed by, for example, Weir (2005), might further enhance any results which an evaluation based on the model produces. This might, however, also limit the usefulness of the model since it could put its application beyond the skills and time availability of EAP practitioners. Thus, despite the limitations of this study as outlined above, there is potential for the study to be further expanded, and it has still made contributions to the general field of EAP assessment, which I will discuss below.

A further area which could be considered an issue is the fact that I only considered staff and student opinions of the assessments being investigated; no other forms of data were evaluated. Although a consideration of student performance data on progression to their main programmes of study could enhance the discussion, this does need to be done with caution since these data are a result of more than just language proficiency.

7.5 Summary of the contributions made

The realities of EAP mean that "language testers are not a separate species from language teachers [and] the vast majority of people who design, prepare, administer and mark language tests are teachers" (Hamp-Lyons, 2000, p. 580). Additionally, in terms of assessment development it is rare for these language test designers to be starting from scratch; rather they are working with assessments which already exist. Before deciding whether changes need to be made to assessments, let alone how this should be done, a good starting point is establishing the usefulness of current assessments. This brings me to the contributions made by this thesis, which I will consider from three key aspects: contributions to theory, contributions to practice and contributions to policy.

Starting with theory, what this study has done is present an adapted practical usefulness framework which can be applied to assessments currently in use and which can be used to inform changes and amendments to these assessments. There are more comprehensive frameworks which exist for validity studies; there are more technical and statistical approaches that could be used to approach evaluating test reliability, but that was never the purpose of this study. The purpose of this study was to provide a framework which can be used, with relative ease, by EAP professionals in order to investigate the usefulness of a battery of EAP assessments in use at a specific institution. A reliability study, as is explained in detail by Weir (2005), implies

complex statistical analysis which may well require a skills-base that is not available to an English language teaching team; an in-depth validity study is a complex task, and one which can take may different forms, but the reality in many English language departments is that there is little time for this level of test development when it comes to in-house tests. Ultimately, then, in terms of theory, the contribution that this study has made to the field is to offer an adapted form of a practical usefulness framework for specific application to in-use, in-house EAP assessments. As a theoretically-derived descriptive framework, I would argue that it can be used by language teaching specialists across the HE sector working in this field.

To consider now the contributions to practice, the results of this thesis have had an impact in three ways. Firstly, they have had a significant impact on the practice of EAP assessment at MUC. As a result of the research carried out for this study, there have been recent changes made to the assessments, leading to greater uniformity across the assessments used on both the IEP and the SEP. Additionally, there have been changes regarding the contents of the assessments. Where assessments were seen to be lacking, changes have been made to improve them, with further research now being carried out to investigate the impact of these changes. Further contributions to practice have been in the form of wider institutional practice. In carrying out a full assessment audit across the institution, awareness of inconsistencies and issues with assessment terminology across MUC have been raised, resulting in a complete overhaul of the approach to assessment design in the institution.

In addition to this, there has been significant impact on the students. In the process of evaluating the in-use assessments, a number of issues were brought to the fore, as is evidenced by the large list of recommendations presented in section 7.3 above. As a

result of this study, the changes recommended have been implemented. This has also had an impact on the pre-sessional EAP programmes' curriculum design, thus leading to a suite of programmes which better prepare students for their life in UK HE. Additionally, the newly-adapted assessments provide greater reliability in terms of the mark schemes and grading criteria, thus ensuring that students are awarded grades which more accurately reflect their ability. Further, by providing recommendations for improvements in terms of the construct representation, the assessments as re-developed on the basis of the recommendations of this study now test more effectively the academic language ability of the students. Finally, by resulting in a more authentic representation of the TLU situations at MUC, the students are being provided with greater developmental opportunities.

Finally, it is at the institutional policy level where this study has had the greatest impact. The first policy implication of the study has been in the form of the measures used to establish the level of the task contents, in particular for the listening and reading exams. As it explained in the discussion chapters, institutional policy to date has been to use Compleat Lexical Tutor as a tool for establishing text difficulty, a tool which makes detailed use of Coxhead's (2000) AWL. The crude measure that this offers, however, has brought this choice into question, and more recently, as a result of additional research, the use of Gardener and Davies' (2014) Academic Vocabulary List and Ackerman and Chen's (2013) Academic Collocation List in conjunction with the AWL has been implemented. Essentially, what this study has contributed to MUC's institutional policy is a document which makes it possible for the English team to further develop policies regarding entry requirements onto the various programmes of study at the university. In basing any future policy changes on a theoretically-supported, empirically-researched study, the university can be more confident that the

decisions that are made on the basis of assessments developed within the institution are useful predictions of how well the students will be able to cope with the demands of their academic studies at MUC. In investigating in detail the assessments, and instigating changes deemed necessary as a result of this investigation, it has made it possible for the English development centre staff to 'defend the indefensible' to students and subject specialist staff with a clear theoretical underpinning for the results.

REFERENCES

Ackerman, K. & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12 (4), 235 – 247. DOI: 10.1016/j.jeap2013.08.002

Alderson, J. C. & Wall, D. (1993). Does Washback Exist? *Applied Linguistics,* 14 (2), 115 – 129. DOI: 10.1093/applin/14.2.115

Al Hajr, F. (2014). The Predictive Validity of Language Assessment in a Pre-University Programme in two Colleges in Oman: A Mixed Method Analysis. *International Journal of Contemporary Research*, 2 (2), 121 – 147. Available from http://imjcr.com/journals/imjcr/Vol_2_No_2_June_2014/8.pdf [Accessed 1 March 2015]

Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal Research in International Education*, 5 (2), 131 – 154. DOI: 10.1177/1475240906065589

Andrade, M. S. (2010). Increasing accountability. Faculty perspectives on the English language competence of nonnative English speakers. *Journal of Studies in International Education*, 14 (3), 221 – 239. DOI: 10.1177/1028315308331295

Aryadoust, V. (2013). *Building a Validity Argument for a Listening Test of Academic Proficiency*. Newcastle-upon-Tyne, UK: Cambridge Scholars Publishing.

Bacha, N. N. (2010). Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes*, 9, 229 – 241. DOI: 10.1016/j.jeap.2010.05.001

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press

Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Baez, B. (2002). Confidentiality in qualitative research: reflections on secrets, power and agency. *Qualitative Research*, 2 (1), 35 – 58.

Baiardi, J. M., Gultekin, L. & Brush, B. M. (2015) Using Sociograms to Enhance Power and Voice in Focus Groups *Public Health Nursing*, 32 (5) DOI: 10.1111/phn.12199

Bailey, K. M. (1999). *Washback in Language Testing*. New Jersey, USA: English Testing Service.

Bailey, A. L. & Huang, B. H. (2011). Do current English language development/ proficiency standards reflect the English needed for success in school? *Language Testing*, 28 (3), 343 - 365. DOI: 10.1177/0265532211404187

Bang, H & Montgomery, D. (2013). Understanding International Graduate Students' Acculturation Using Q Methodology. *Journal of College Student Development,* 54 (4), 343 – 360. DOI: 10.1353/csd.2013.0066

Bannerjee, J. & Wall, D. (2006). Assessing and reporting performances on pre-sessional EAP course: Developing a final assessment checklist and investigating its validity. *Journal of English for Academic Purposes*, 5, 50 – 69. DOI: 10.1016/j.jeap.2005.11.003

Barnett, R. (2007). Assessment in higher education: an *impossible mission*?. In D. Boud and N. Falichikov (Eds.). *Rethinking Assessment in Higher Education: Learning for the longer term*. (pp. 29 – 40). Oxford, UK: Routledge.

Bassey, M. (1981). Pedagogic Research: on the relative merits of search for generalisation and the study of single events. *Oxford Review of Education*, 7 (1), 73 – 94. DOI: 10.1080/0305498810070108

Basturkmen, H. (2002). Negotiating meaning in seminar-type discussion and EAP. *English for Specific Purposes*, 21 (3), 233 – 242. DOI: 10.1016/S0889-4906(01)00024-2

Bejar, I., Douglas, D., Jamieson, J., Nissan, S. & Turner, J. (2000). *TOEFL 2000 Listening Framework: A Working Paper*. New Jersey, USA: Educational Testing Service

Bell, J. (2005). *Doing Your Research Project: A Guide for First-Time Researchers in Education, Health and Social Science*. 4th Edition. Maidenhead, UK: McGraw-Hill

Bell, L. & Nutt, L (2002) Divided Loyalties, Divided Expectations: Research Ethics, Professional and Occupational Responsibilities, in M. Mauthner , M. Birch , J. Jessop and T. Miller (eds) *Ethics in Qualitative Research*, pp. 70-90. London: Sage

Benesch, S. (2001). *Critical English for Academic Purposes: Theory, Politics and Practice*. New York, USA: Routledge

Benson, M. J. (1994). Lecture listening in an ethnographic perspective. In Flowerdew, J. (Ed.). *Academic English: research perspectives*. Cambridge: Cambridge University Press. 181-198.

Bergman, L. (2016) Supporting academic literacies: university teachers in collaboration for change *Teaching in Higher Education* 21 (5), pp. 516 – 531. DOI: 10/1080/13562517.1160220

Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning *Assessment in Education*, 5 (1), 7 – 74.

Black, P. & Wiliam, D. (2003). 'In Praise of Educational Research': formative assessment. *British Educational Research Journal*, 29 (5), 623 – 637. DOI: 10.1080/0141192032000133721

Bloxham, S. & Boyd, P. (2007). *Developing Effective Assessment in Higher Education: A Practical Guide*. Maidenhead, UK: The Open University Press

Bonanno, H. & Jones, J. (2007). *The MASUS Procedure: Measuring the Academic Skills of University Students A Diagnostic Assessment*. Sydney, Australia: The Learning Centre, The University of Sydney.

Boud, D. & Falichikov, N. (2007). Introduction: Assessment for the longer term. In D. Boud and N. Falichikov (Eds.). *Rethinking Assessment in Higher Education: Learning for the longer term*. (pp. 3 – 13). Oxford, UK: Routledge.

Boud, D., Cohen, R. & Sampson, J. (1999). Peer Learning and Assessment. *Assessment & Evaluation in Higher Education*, 24 (4), 413 – 426. DOI: 10.1080/0260293990240405

Bravo-Moreno, A. (2003). Power Games between the Researcher and the Participant in Social Enquiry. *The Qualitative Report*, 8 (4), 624 – 639. Available from www.nova.edu/ssss/QR/QR8-4/bravo.pdf [Accessed 23 May 2017]

Brennan, R. L. (2013). Commentary on "Validating the Interpretations and Uses of Test Scores", *Journal of Educational Measurement*, 50 (1), 74 – 83

Bridgeman, M. & Carlson, S. B. (1984). Survey of Academic Writing Tasks. *Written Communication*, 1 (2), 247 – 280. DOI: 10.1177/0741088384001002004

Brinkman, S. & Kvale, S. (2005). Confronting the ethics of qualitative research. *Journal of Constructivist Psychology*, 18, 157 – 181. DOI: 10.1080.10720530590914789

Brown, S. & Knight, P. (1994). *Assessing Learners in Higher Education.* Abingdon, UK: Routledge

Bryan, C. & Clegg, K. (2006). *Innovative Assessment in Higher Education*. Oxford, UK: Routledge.

Butler, F. A., Eignor, D., Jones, S., McNamara, T. & Suomi, B. K. (2000). *TOEFL 2000 Speaking Framework: A Working Paper*. New Jersey, USA: Educational Testing Service.

Bygrave, C., Asik-Dzidar, O. & Saini, G. K. (2014). From homogeneity to heterogeneity: the multicultural future of higher education. *On the Horizon*, 22 (3), 199 – 209. DOI: 10.1108/OTH-05-2014-0016

Campbell, A. & Norton, L. (Eds.). *Learning, Teaching and Assessing in Higher Education: Developing Reflective Practice*. Exeter, UK: Learning Matters.

Canale, M. & Swain, M (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1 (1), 1 – 47. DOI: 10.1093/applin/I.1.1

Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment and Evaluation in Higher Education*, 34 (1), 79 – 89. DOI: 10.1080/02602930801895786

Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences, *Language Testing*, 33 (4), 453 – 472. DOI: 10.1177/0265532215593312

Chanock, K. (2010) The right to reticence, *Teaching in Higher Education*, 15 (5), pp. 543 – 552. DOI: 10.1080/13562517.2010.491907

Cheng, L. & Curtis, A. (2010). *English Language Assessment and the Chinese Learner*. New York, USA: Routledge.

Cheng, R. & Erben, A. (2012). Language Anxiety: Experiences of Chinese Graduate Students at U.S. Higher Institutions. *Journal of Studies in International Education*, 16 (5), 477-497. DOI: 10.1177/1028315311421841

Christison, M. A. & Krahnke, K. J. (1986). Student Perceptions of Academic Language Study. *TESOL Quarterly*, 20 (1). 61 – 81. Available from http://www.jstor.org/stable/3586839 [Accessed 15 October 2013]

Clegg K. & Bryan C. (2006). Reflections, rationales and realities. In C. Bryan & K. Clegg. *Innovative Assessment in Higher Education*. (pp. 216 – 227). Oxford, UK: Routledge.

Clements, M. D. & Cord, B. A. (2013). Assessment guiding learning: developing graduate qualities in an experiential learning programme. *Assessment & Evaluation in Higher Education*, 38 (1), 114 – 124. DOI: 10.1080/02602938.2011.609314

Clifford, J. (2012). *The Importance of Researcher Positionality: A discussion of how positionality affects the research process for educational researchers* . (Ed.D Assignment 3, University of Sheffield, Sheffield, United Kingdom).

Clough, P. & Nutbrown, C. (2012). *A Student's Guide to Methodology*. London, UK: Sage Publications.

Clovis, C. (2007) Re-conceptualizing Assessment at the University Level. In K. Smith (Ed.). *Teaching, Learning, Assessing: A Guide for Effective Teaching at College and University*. (pp. 137 – 142). Oakville, Canada: Mosaic Press.

Colucci, E. (2007). "Focus Groups Can Be Fun": The Use of Activity-Oriented Questions in Focus Group Discussions. *Qualitative Health Research,* 17 (10), 1422 – 1433. DOI: 10.1177/1049732307308129

Corbin, J. & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA, USA : Sage Publications.

Cordon, A. & Sainsbury, R. (2006). Exploring 'Quality': Research Participants' Perspectives on Verbatim Quotations. *International Journal of Social Research Methodology*, 9 (2), 97 – 110. DOI: 10.1080/13645570600595264

Cottrell, S. (2013). *The Study Skills Handbook.,* 4th Edition. Basingstoke, UK: Palgrave Macmillan

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Cambridge University Press. Available from: http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf [Accessed 1 August 2012]

Coxhead, A. J. (1998). *An academic word list.* (English Language Occasional Publication No. 18). Wellington, New Zealand: University of Wellington.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34 (2), 213 – 238. DOI: 10.2307/3587951

Coxhead, A. (2011). The Academic Word List 10 Years On: Research and Teaching Implications. *TESOL Quarterly*, 45 (2), 355 – 362.

Crisp, G. T. (2012). Integrative assessment: reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, 37 (1), 33 – 43. DOI: 10.1080/02602938.2010.494234

Cumming, A., Kantor, R., Powers, D., Santos, T. & Taylor, C. (2000). *TOEFL 2000 Writing Framework: A Working Paper*. New Jersey, USA: Educational Testing Service.

Davies, A. (Ed.). (1968). *Language Testing Symposium*. Oxford, UK: Oxford University Press.

Davies, A. (1977). The construction of language tests. In J. B. P. Allen & A. Davies (Eds.). *The Edinburgh Course in Applied Linguistics Volume 4: Testing and Experimental Methods*. Oxford, UK: Oxford University Press.

Davies, A. (1990). *Principles of Language Testing*. Oxford, UK: Blackwell Publishing.

Denscombe, M. (2008). *The Good Research Guide for Small Scale Research Projects*, 3rd Edition. Maidenhead, UK: The Open University Press.

Denzin, N. K. & Lincoln, Y. S. (Eds.). (2003). *Collecting and Interpreting Qualitative Materials*. 2nd Edition. London, UK: Sage

Denzin, N. K. & Lincoln, Y. S. (Eds.). (2008). *Collecting and Interpreting Qualitative Materials*. 3rd Edition. London, UK: Sage

Deutsch, N. (2004). Positionality and the Pen: Reflections on the Process of Becoming a Feminist Researcher and Writer. *Qualitative Enquiry*, 10 (6), 885 – 902. DOI: 10.1177/1077800404265723

Donohue, J. P. & Erling, E. J. (2012). Investigating the relationship between the use of English for academic purposes and academic attainment. *Journal of English for Academic Purposes*, 11 (2), 210 – 219. DOI: 10.1016/j.jeap.2012.04.003

Dooey, P. (2010). Students' Perspectives of an EAP Pathway Program. *Journal of English for Academic Purposes*, 9 (3), 184-197. DOI: 10.1016/j.jeap.2010.02.013

Douglas, D. (2001). Language for Specific Purposes assessment criteria: where do they come from? *Language Testing*, 18, 171 – 185. DOI: 10.1177/206553220101800204

Douglas, D. (2010). *Understanding Language Testing*. London, UK: Hodder Education.

Dunn, L., Morgan, C., O'Reilly, M. & Parry, S. (2004). *The Student Assessment Handbook*: *New Directions in Traditional & Online Assessment*. London, UK: RoutledgeFalmer.

Durkin, K. (2008). The adaptation of East Asian masters students to western norms of critical thinking and argumentation in the UK. *Intercultural Eduation*, 19 (1), pp. 15 – 27. DOI: 10.1080/14675980701852228

Dzakiria, H. (2012). Theory of Relatability as a Possible Alternative to the Issue of Generalising of Research Findings: The Case of Open and Distance Learning (ODL) at Universiti Utara Malaysia. *Malaysian Journal of Distance Education*, 14 (1), 41 – 58. Available from: http://www.lib.uum.edu.my:8080/bitstream/123456789/1153/1/1092.pdf [Accessed 23 October 2016]

Easton, K. L., Fry McComish, J. & Greenberg, R. (2000). Avoiding Common Pitfalls in Qualitative Data Collection and Transcription. *Qualitative Health Research,* 10 (5), 703 – 707.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K. & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *SAGE Open*, 4 (1), 1 – 10. DOI: 10.1177/2158244014522633

Elwood, S. A. & Martin, D. G. (2000). "Placing" Interviews: Location and Scales of Power in Qualitative Research. *The Professional Geographer*, 52 (4), 649 – 657. DOI: 10.1111/0033-0124.00253

Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P. & Schedl, M. (2000). *TOEFL 2000 Reading Framework: A Working Paper*. New Jersey, USA: Educational Testing Service.

Estaji, M. & Tajeddin, Z. (2012). The Learner Factor in Washback Context: An Empirical Study Investigating the Washback of the IELTS Academic Writing Test. *Language Testing in Asia*, 2 (1), 5 – 25.

Evans, S. & Morrison, B. (2011). The first term at university: implications for EAP. *ELT Journal*, 65 (4), 387 – 397. DOI: 10.1093/elt/ccq0972

Ferris, D. & Tagg, T. (1996a). Academic Oral Communication Needs of EAP Learners: What Subject-Matter Instructors Actually Require. *TESOL Quarterly*, 30 (1), 31 – 58. Available from: http://www.jstor.org/stable/3587606 [Accessed 15 October 2013]

Ferris, D. & Tagg, T. (1996b). Academic Listening/Speaking Tasks for ESL Students: Problems, Suggestions and Implications. *TESOL Quarterly*, 30 (2), 297 – 320. Availble from: http://www.jstor.org/stable/3588145 [Accessed 15 October 2013]

Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P. & Rees, M. (2012). Faculty and Students Conceptions of Assessment in Higher Education. *Higher Education,* 64 (1), 119 – 133. DOI: 10.1007/s10734-011-9484-1

Flowerdew, J. & Miller, L. (1995). On the Notion of Culture in L2 Lectures. *TESOL Quarterly*, 29 (2), 345 – 373. DOI: 10.2307/3584628

Forbat, L & Henderson, J. (2005). Theoretical and Practical Reflections on Sharing Transcripts With Participants. *Qualitative Health Research*, 15 (8), 1114 – 1128. DOI: 10.1177/1049732305279065

Franz, N. K. (2011). The Unfocused Focus Group: Benefit or Bane? *The Qualitative Report*, 16 (5), 1380 – 1388.

Fulcher, G. (2010). *Practical Language Testing*. London, UK: Hodder Education.

Fulcher, G. & Davidson, F. (2007). *Language Testing and Assessment: An advanced resource book*. Abingdon, UK: Taylor and Francis

Gardner, D & Davies, M (2014). A New Academic Vocabulary List. *Applied Linguistics*, 35 (3), 305 – 327. DOI: 10.1093/applin/amt015

Gibbs, G. (2006). How assessment frames student learning. In C. Bryan & K. Clegg. *Innovative Assessment in Higher Education*. (pp. 23 – 36). Oxford, UK: Routledge.

Giroux, H. (1988). *Teachers as intellectuals: Toward a critical pedagogy of learning*. Westport, USA. Bergin and Garvey.

Goldingay, S., Hitch, D., Carrington, A., Nipperness, S. & Rosario, V. (2016) Transforming roles to support student development of academic literacies: a reflection on one team's experience. *Reflective Practice*, 17 (3) pp. 334 – 346. DOI: 10.1080/14623943.2016.1164682

Grabe, W. & Zhang, C. (2013) Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL Journal*, 4 (1), pp. 9 – 24.

Graneheim, U. H. & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nursing Education Today,* 24, 105 – 112. DOI: 10.1016/j.nedt.2003.10.001

Green, A. (2013). Washback in language testing. *International Journal of English Studies*, 13 (2), 39 – 51.

Guse, J. (2011). *Communicative Activities for EAP*. Cambridge, UK: Cambridge University Press.

Ha P. L. (2006). Plagiarism and overseas students: Stereotypes again? *ELT Journal* 60 (1), 76–78.

Hager, P. & Butler, J. (1996). Two Models of Educational Assessment. *Assessment & Evaluation in Higher Education*, 21 (4), 367 – 378. DOI: 10.1080/0260293960210407

Hallett, F (2013) Study support and the development of academic literacy in higher education: a phenomenographic analysis, *Teaching in Higher Education*, 18 (5) pp. 518 – 530 DOI: 10.1080/13562517.2012.752725

Hammesley-Fletcher, L. & Hanley, C. (2016). The use of critical thinking in higher education in relation to the international student: Shifting policy and practice. *British Education Research Journal*, 42 (6), 978 – 992. DOI: 10.1002/berj.3246

Hamp-Lyons, L. (2000). Social, Professional and Individual Responsibility in Language Testing. *System*, 28 (4), 579 – 591.

Hamp-Lyons, L. (2001). English for academic purposes. In R. Carter, D. Nunan (Eds.), The Cambridge guide to teaching English to speakers of other languages, Cambridge, UK: Cambridge University Press.

Harmer, J. (2007). *The practice of English language teaching*. 4th edition. Harlow: Longman.

Hatipoğlu, Ç. (2010) Summative Evaluation of an English Language Testing and Evaluation Course for Future English Language Teachers in Turkey. *ELTED*, 13, 40 – 51. Available from: http://www.elted.net/issues/volume-13/Vol%2013%20Paper%205%20Hatipoglu.pdf [Accessed 15 October 2013]

Heaton, J. B. (1975). *Writing English Language Tests*. Harlow, UK: Longman.

Hefce (2016). National Student Survey. Available from: http://www.hefce.ac.uk/lt/nss/ [Accessed 3rd June 2017]

Hernández, R. (2012). Does Continuous Assessment in Higher Education Support Student Learning? *Higher Education*, 64 (4), 489 – 502. DOI: 10.1007/s10734-012-9506-7

Herod, A. (1999), Reflections on interviewing foreign elites: praxis, positionality, validity, and the cult of the insider. *Geoforum,* 30 (4) 313-327. DOI: 10.1016/S0016-7185(99)00024-X

Holder, G. M., Jones, J., Robinson, R. A. & Krass, I. (1999). Academic Literacy Skills and Progression Rates Amongst Pharmacy Students. *Higher Education Research & Development*, 18 (1), 19 – 30. DOI: 10.1080/0729436990180103

Home Office. (2016). *Tier 4 of the Points Based System – Policy Guidance*. London, UK: Home Office.

Hughes, A. (2003). *Testing for Language Teachers*, 2nd Edition. Cambridge, UK: Cambridge University Press

Hughes ,C. (2013). A case study of assessment of graduate learning outcomes at the programme, course and task level. *Assessment & Evaluation in Higher Education*, 38 (4), 492 – 506. DOI: 10.1080/02602638.2012.658020

Hyland, K. & Hamp-Lyons, L. (2002). EAP: issues and directions. *Journal of English for Academic Purposes*, 1 (1), 1 – 12. DOI: 10.1016/S1475-1585(02)0002-4

Hyland, K. & Tse, P. (2007). Is there an "Academic Word List"? *TESOL Quarterly*, 41 (2), 235 – 253. DOI: 10.1002/j.1545-7249.2007.tb00058.x

IELTS. (2016). Writing Task 2: Band Descriptors (Public Version). Available from: https://www.ielts.org/~/media/pdfs/writing-band-descriptors-task-2.ashx [Accessed 7 January 2016]

Johns, A. M. (1981). Necessary English: A Faculty Survey. *TESOL Quarterly*, 15 (1), 51 – 57. Available from: http://www.jstor.org/stable/3586373

Jordan, R. R. (1997). *English for Academic Purposes: A Guide and Resource Book for Teachers*. Cambridge, UK: Cambridge University Press

Jordan, R. R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes*, 1, 69 – 78.

Kane, M. T. (2006) Validation.. In R. Brennan (Ed.), *Educational Measurement*, 4th ed. (pp. 17-64), Westport, CT: American Council on Education and Praeger.

Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing,* 29 (1), 3 – 17.

Kane, M. T. (2013a). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement,* 50 (1), 1 – 73.

Kane, M. T. (2013b). Validation as a Pragmatic, Scientific Activity. *Journal of Educational Measurement*, 50 (1), 115 – 122.

Kane, M., Crooks, T. & Cohen, A. (1999). Validating Measures of Performance. *Educational Measurement: Issues and Practice*, 18 (2), 5 – 17. DOI: 10.1111/j.1745-3992.1999.tb00010.x

Karnieli-Miller, O., Strier, R. & Pessach, L. (2009). Power Relations in Qualitative Research. *Qualitative Health Research*, 19 (2), 279 – 289. DOI: 10.1177/1409732308329306

Kember, D. (2000). Misconceptions about the learning approaches, motivation and study practices of Asian students. *Higher Education*, 40 (1), 99-121. DOI: 10.1023/A:1004036826490

Kidd, P. S. & Parshall, M. B. (2000). Getting the focus and the group: enhancing analytical rigor in focus group research. *Qualitative Health Research*, 10 (3), 293 – 308. DOI: http://dx.doi.org/10.1177/104973200129118453

Kiely, R. (2004). Learning to critique in EAP. *Journal of English for Academic Purposes*, 3 (3), 211 – 227. DOI: 10.1016/j.jeap2003.11.004

Kingston, E. & Forland, H. (2008). Bridging the Gap in Expectations between International Students and Academic Staff. *Journal of Studies in International Education*, 12 (2), p.204-221. DOI: 10.1177/1028315307307654

Knight, P. T. (2002). Summative Assessment in Higher Education: Practices in disarray. *Studies in Higher Education*, 27 (3). 275 – 286. DOI: 10.1080/03075070220000662

Kreuger, R. A. (1993). Quality Control in Focus Group Research. In D. L. Morgan (Ed.). *Successful Focus Groups: Advancing the State of the Art*. London, UK: SAGE Publications.

Kumaravadivelu, B. (2003). Problematizing cultural stereotypes in TESOL. *TESOL Quarterly*, *37* (4), 709-718.

Lado, R. (1961). *Language Testing*. London, UK: Longman.

Lau, A. M. S. (2016). 'Formative good, summative bad?' – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, 40 (4), 509 – 525. DOI: 10.1080/0309877X.2014.984600

Lea, M. (1998). Academic literacies and learning in higher education: constructing knowledge through texts and experience. *Studies in the Education of Adults*, 30 (2), 156 – 171. DOI: 10.1080/02660830.1998.11730680

Lea, M. (2004). Academic literacies: a pedagogy for course design. *Studies in Higher Education*, 29 (6), pp. 739 – 756. DOI: 10.1080/ 0307507042000287230

Lea, M. R. & Street, B. V. (1998). Student Writing in Higher Education: an academic literacies approach. *Studies in Higher Education*, 23 (2), 157 – 172. DOI: 10.1080/03075079812331380364

Lebcir, M., Wells, H. & Bond, A. (2008). Factors affecting academic performance of international students in project management courses: a case study from a British Post 92 University. *International Journal of Project Management*, 26 (3). 268 – 274. DOI: 10.1016/j.ijproman.2008.02.003

Lewis, M. & Reinders, H. (2003). *Study skills for speakers of English as a second language*. Basingstoke, UK: PalgraveMacmillan.

Lichtman, M. (2006). *Qualitative Research in Education: A user's guide*. Thousand Oaks, USA: Sage Publications.

Lillis, T. & Scott, M. (2007). Defining academic literacies research: issues of epistemology, ideology and strategy. *Journal of Applied Linguistics*, 4 (1), 5 – 32. DOI: 10.1558//japl.v4il.5

Lillyman, S. & Bennett, C. (2014). Providing a positive learning experience for international students studying at UK universities: A literature review. *Journal of Research in International Education*, 13 (1), 63 – 75. DOI: 10.1177/1475240914529859

Lincoln, D. & Kearney, M-L. (2015). Competence assessment in higher education. *Studies in Higher Education,* 40 (3), 391 – 392. DOI: 10.1080/03075076.2015.1005334

Lincoln, Y & Guba E. (1985). *Naturalistic Inquiry*. Beverley Hills, CA: Sage

Linville, D., Lambert-Shute, J., Fruhauf, C. A., & Piercy, F. P. (2003). Using participatory focus groups of graduate students to improve academic departments: A case example. *The Qualitative Report*, 8 (2), 210-223. Available from: http://www.nova.edu/ssss/QR/QR8-2/linville.pdf [Accessed 14 October 2014]

Lissitz, R. W. & Samuelsen, K. (2007). A Suggested Change in Terminology and Emphasis Regarding Validity and Education. *Educational Researcher,* 36 (8), 437 – 448. DOI: 10.3102/0013189X07311286

Littlewood, W. (2000) Do Asian students really want to listen and obey? *ELT Journal*, 54, pp. 31 – 36

Luna, C. (2002). Learning from diverse learners: (Re)writing academic literacies and learning disabilities in college. *Journal of Adolescent & Adult Literacy*, 45 (7), 596 – 605.

Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18 (351 – 372.

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, 10. 79 – 88. DOI: 10.1016/j.jeap.2011.03.001

Madsen, H. S. (1983). *Techniques in Testing*. New York, USA: Oxford University Press.

Malone, M. E. & Montee, M (2014). *Stakeholders' Beliefs About the TOEFL iBT ® Test as a Measure of Academic Language Ability*. New Jersey, USA: Educational Testing Service.

Martin, R. J. & Van Gunten, D. M. (2002). Reflected Identities: Applying Positinality and Multicultural Social Reconstructionism in Teacher Education. *Journal of Teacher Education*, 53 (1), 44 – 54. DOI: 10.1177/0022487102053001005

Mason, A. (1995). By dint of: Student and lecturer perceptions of lecture comprehension strategies in first-term graduate study. In J. Flowerdew (Ed.), *Academic Listening: Research Perspectives* (pp. 199–218).

Mason, M. (2007) Critical Thinking and Learning, *Educational Philosophy and Theory*, pp. 339 – 349. DOI: 10.1111/j.1469-5812-2007.00343.x

McMillan, K. & Weyes, J. (2012) *The Study Skills Book*, 3rd Edition. Harlow, UK: Person Education

McMurray, P. (2011). *Study Skills Essentials: Oxford graduates reveal their study tactics, essay secrets and exam advice*. Hollywood, UK: Effective Study Skills Publications

McNamara, T. F. (1996). *Measuring second language performance*. London and New York: Longman.

McNamara, T. (2000). *Language Testing*. Oxford, UK: Oxford University Press.

McNamara, T. (2006). Validity in Language Testing: The Challenge of Sam Messick's Legacy. *Language Assessment Quarterly*, 3 (1), 31 – 51. DOI: 10.1207/s15343411laq0301_3

McNamara, T. & Roever, C. (2006). *Language Testing: The Social Dimension*. Oxford, UK: Blackwell.

Messick, S. (1989). 'Validity' in R. L. Linn (ed.) *Educational Measurement.* 3rd Edition. (pp. 13 – 103). London, UK: Macmillan.

Messick, S. (1996). *Validity and Washback in Language Testing*. New Jersey, USA: Educational Testing Service.

Miles, M. B. & Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd ed. London, UK: SAGE

Miller, T. & Bell, L. (2002) Consenting to what? Issues of access, gate-keeping and 'informed' consent. In Mauthner, M., Birch, M., Jessop, J. & Miller, T. (2002) *Ethics in Qualitative Research* London, UK: Sage.

Moore, S., Neville, C., Murphy, M. & Connolly, C. (2010). *The Ultimate Study Skills Handbook*. Maidenhead, UK: Open University Press

Morgan, D. L. & Kreuger, R. A. (1993).When to use focus groups and why. In D. L. Morgan (Ed.). *Successful Focus Groups: Advancing the State of the Art*. London, UK: SAGE Publications.

Morgan, L. (1996). Focus Groups. *Annual Review of Psychology*, 22, 129 – 152.

Moss, P. A. (2007). Reconstructing Validity. *Educational Researcher,* 36 (8), 470 - 476. DOI: 10.3102/0013189X07311608

Mullings, B. (1999), Insider or outsider, both or neither: some dilemmas of interviewing in a cross-cultural setting.
Geoforum, 30 (4), 337-350. DOI: 10.1016/S0016-7185(99)00025-1

Newton, P. E. (2013). Two Kinds of Argument? *Journal of Educational Management*, 50 (1), 105 – 109. DOI: 10.1111/jedm.12004

Norton, L. (2007). Using assessment to promote quality learning in higher education. In A. Campbell and L. Norton (Eds.). *Learning, Teaching and Assessing in Higher Education: Developing Reflective Practice*. (pp. 92 – 101). Exeter, UK: Learning Matters.

O'Donoghue, T. (2007). *Planning Your Qualitative Research Project: An introduction to interpretive research in education*. Abingdon, UK: Routledge

O'Leary Z (2004). *The essential guide to doing research*. London: Sage Publications

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30 (3), 363 – 380. DOI: 10.1177/0265532213480336

Opie, C. (2004). (Ed.) *Doing Educational Research: A Guide to First Time Researchers*. London, UK: SAGE Publications.

Ostler, S. E. (1980). A Survey of Academic Needs for Advanced ESL. *TESOL Quarterly*, 14 (4), 489 – 502. Available from http://www.jstor.org/stable/3586237 [Accessed 17 October 2013]

Pereira, D., Assunção, F. & Niklasson, L. (2016). Assessment revisited: a review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*, 41 (7), 1008 – 1032. DOI: 10.1080/02602938.2015.1055233

Phakiti, A., Hirsch, D. & Woodrow, L. (2013). It's not only English: Effects of other individual factors on English language learning and academic learning of ESL international students in Australia. *Journal of Research in International Education*, 12 (3), 239 – 258. DOI: 10.1177/1475240913513520

Pierson, H. D. (1996) Learner culture and learner autonomy in the Hong Kong Chinese context in R. Pemberton *et al*. (Eds.) *Taking control: Autonomy in language learning*, Hong Kong: Hong Kong University Press, pp. 49 - 58

Pilliner, A. (1968). Subjective and objective testing. In A. Davies (Ed.). *Language Testing Symposium: A Psycholinguistic Approach*. (pp. 19 – 35). Oxford, UK: Oxford University Press.

Pillow, W. (2003). Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *International Journal of Qualitative Studies in Education,* 16 (2), 175 – 196.
DOI: 10.1080/0951839032000060635

Popham, W. J. (1997). Consequential Validity: Right Concern – Wrong Concept, *Educational Measurement: Issues and Practice*, 16 (2), 9 – 13. DOI: 10.1111/j.1745-3992.1997.tb00586.x

Pot, A. & Weideman, A. (2015). Diagnosing academic language ability: An analysis of the Test of Academic Literacy for Postgraduate Students. *Language Matters*, 46 (1), 22 – 43. DOI: 10.1080/10228195.2014.986665

Price, M. & Donovan, B. (2006). Improving performance through enhancing student understanding of criteria and feedback. In C. Bryan & K. Clegg. *Innovative Assessment in Higher Education*. (pp. 100 – 109). Oxford, UK: Routledge.

Price, M., Carroll, J., O'Donovan, B & Rust, C. (2011). If I was going there I wouldn't start from here: a crucial commentary on current assessment practice. *Assessment & Evaluation in Higher Education*, 36 (4), 479 – 492. DOI: 10.1080/02601930903512883

Qi, L. (2011). A review of washback studies. *Foreign Language Learning Theory and Practice*, 33 (4), 23 – 28.

Ramachandarn, N. T. (2011). Enhancing international students' experiences: An imperative agenda for universities in the UK. *Journal of Research in International Education*, 10 (2), 201 – 220. DOI: 10.1177/1475240911413204

Read, J. (2002). The use of interactive input in EAP listening assessment. *Journal of English for Academic Purposes*, 1, 105 – 119.

Rossman, G. B. & Rallis, S. F. (2003). *Learning in the Field: An Introduction to Qualitative Research*. 2nd Edition. London

Ryan, J. & Louie, K. (2007) False Dichotomy? 'Western' and 'Confucian' concepts of scholarship and learning. *Education Philosophy and Theory*, DOI: 10.1111/j.169-5812.2007.00347x

Ryan, K., Gandha, T., Culbertson, C. & Carlson, C. (2013). Focus Group Evidence: Implications for Design and Analysis. *American Journal of Evaluation,* 35 (3), 328 - 345, DOI: 10.1177/1098214013508300

Schmitt, D. & Hamp-Lyons, L. (2015). The need for EAP teacher knowledge in assessment. *Journal of English for Academic Purposes*, 18, 3 – 8.
DOI: 10.1016/j.jeap.2015.04.003

Scott, D. & Morrison, M. (2007). *Key Ideas in Educational Research*. London, UK: Continuum

Shepherd, L. A. (1997). The Centrality of Test Use and Consequences for Test Validity, *Educational Measurement: Issues and Practice*, 16 (2), 5 – 24. DOI: 10.1111/j.1745-3992.1997.tb00585.x

Shi, L. (2012). Rewriting and paraphrasing source texts in second language writing. *Journal of Second Language Writing*, 21 (2), 134 – 148.
DOI: 10.1016/j.jslw.2012.03.003

Shrodike, A., Hocking, C., Bunrayong, W., Vittayakorn, S., Rattakorn, P., Pierce, D, & Wright-St Clair, V. A. (2017). Research as relationship: engaging with ethical intent. *International Journal of Social Research Methdology*, 20 (3), 285 – 298.
DOI: 10.1080/13645579.2017.1287874

Sigurðsson, G. (2017) Transformative Critique: What Confucianism Can Contribute to Contemporary Education. *Studies in Philosophical Education*, 36, pp. 131 – 146. DOI: 10.1007/s11217-015-9502-3

Sikes, P. (2004). Methodology, Procedures and Ethical Concerns. In C. Opie (Ed.). *Doing Educational Research: A Guide to First Time Researchers*. (pp. 15 – 33). London, England: SAGE Publications.

Sim, J. (1998). Collecting and analysing qualitative data: issues raised by the focus group. *Journal of Advanced Nursing*, 28 (2), 345 – 352. DOI: 10.1046/j.1365-2648.1998.00692.x

Sireci, S. (2007). On Validity Theory and Test Validation. *Educational Researcher,* 36 (8), 477 – 481. DOI: 10.3102/0013189X07311609

Sireci, S. G. (2013). Agreeingon Validity Arguments. *Journal of Educational Management*, 50 (1), 99 – 104. DOI: 10.1111/jedm.12005

Skyrme, G. (2010). Is this a stupid question? International undergraduate students seeking help from teachers during office hours. *Journal of English for Academic Purposes*, 9 (3), 211 – 221. DOI: 10.1016/j.jeap.2010.02.012

Smith, K. E. (2006). Problematising power relations in 'elite' interviews. *Geoforum*, 37, 643 – 653. DOI: 10.1016/j.geoforum.2005.11.002

Smith, M. W. (1995) Ethics in Focus Groups, *Qualitative Health Research*, 5 (4), pp. 478 – 486

Smith, R. A. & Khawaja, N. G. (2010). A review of the acculturation experiences of international students. *International Journal of Intercultural Relations*, 35 (6), 699-713. DOI: 10.1016/j.ijintrel.2011.08.004

Smithson, J. (2000). Using and analysing focus groups: limitations and possibilities. *International Journal of Social Research Methodology*, 3 (2), 103 – 119. Available from: http://www.sfu.ca/~palys/Smithson-2000-Using&AnalysingFocusGroups.pdf [Accessed 15 October 2013]

Song, M. Y. (2011). Note-taking quality and performance on an L2 academic listening test. *Language Testing*, 29 (1), 67 – 89. DOI: 10.1177/026553221415379

Spector-Cohen, E., Kirschner, M. & Wexler, C. (2001). Designing EAP reading courses at the university level. *English for Specific Purposes*, 20 (4), 367 – 386. DOI: 10.1016/S0889-4906(00)00019-3

Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed.) (pp. 443–466). Thousand Oaks, USA: SAGE.

Suskie, L. (2009). *Assessing Student Learning: A Common Sense Guide*. 2nd Edition. San Francisco, USA: Jossey-Bass [Kindle Edition]

Tajino, A., James, R. & Kijima, K. (2005). Beyond needs analysis: soft systems methodology for meaningful collaboration in EAP course design. *Journal of English for Academic Purposes*, 4 (1), 27 – 42. DOI: 10.1016/j.jeap.2004.01.001

Taras, M. (2005). Assessment: Summative and Formative: Some Theoretical Reflections. *British Journal of Educational Studies*, 53 (4), 466 – 478.

Taras, M. (2008). Assessment for learning: sectarian divisions of terminology and concepts. *Journal of Further and Higher Education*, 32 (4), 389 – 397. DOI: 10.1080/030987702395892

Taylor, L. & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes*, 10, 89 – 101. DOI: 10.1016/j.jeap.2011.03.002

Thomas, G. (2009). *How to do Your Research Project: A guide for students in education and applied social sciences.* London, England: SAGE Publications.

Turner, J. (2004). Language as academic purpose. *Journal of English for Academic Purposes*, 3 (2), 95 – 109. DOI: 10.1016/S1475-1585(03)00054-7

Vaismoradi, M., Bondas, T. & Turunen, H. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences*, 15 (3), 398 – 405. DOI: 10.1111/nhs.12048

Wakeford, R. (2003). Principles of student assessment. In H. Fry, S. Ketteridge & S. Marshall (Eds.). *A Handbook for Teaching & Learning in Higher Education* (2nd ed.) (pp. 42 – 61). London, UK: Kogan Page.

Walklin, L. (1990). *Teaching and Learning in Further and Adult Education*. London, UK: Nelson Thornes.

Watson Todd, R. (2003). EAP or TEAP? *Journal of English for Academic Purposes,* 2, 147 – 156. DOI: 10.1016/S1475-1585(03)00014-6

Weigle, S. C. & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21 (2), 118 – 133. DOI: 10.1016/j.jslw2012.03.004

Weir, C. J. (2005) *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke, UK: Palgrave Macmillan.

Weir, C. J. & Wu, J. R. W. (2006) Establishing test form and individual task comparability: a case study of a semi-direct speaking test. *Language Testing*, 23 (2), 167 – 197. DOI: 10.1191/0265532206lt326oa

Wellington, J. (2000) *Educational Research: Contemporary Issues and Practical Approaches*. London UK: Continuum

Wellington J., Bathmaker A., Hunt C., McCulloch G. & Sikes P. (2005). *Succeeding with your doctorate*. London UK: Sage.

Wilkinson, S. (2011) How Useful Are Focus Groups in Feminist Research? In R. S. Barbour & J Kitzinger (Eds.) (2011) *Developing Focus Group Research*, SAGE pp. 64 - 78 DOI: http://dx.doi.org/10/ 4135/9781849208857 [accessed 29/04/2017]

Wong, F. F. & Thang, S. M. (2008). Developing Academic Competency for Studies in English: The Malaysian ESL Teachers' Perspective. *English for Specific Purposes World*, 4 (1), 1 – 28. Retrieved from: http://www.ukm.my/smthang/pdf/2008_1.pdf 15/07/2015

Wu, G. C. H. & Hung, D. C. H. (2011). Exploring a Teacher's Beliefs in an EAP Course: A Qualitative Perspective. *Taiwan International EAP Journal*, 3 (1), 31 – 64.

Xi, X. (2008). Methods of Test Validation. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (pp. 2316-2335). Boston, MA: Springer US.

Ypsilandis, G. S. & Kantaridou, Z. (2007). English for Academic Purposes: Case Studies in Europe. *Revista de Lingüística y Lenguas Aplicadas*, 2, 69 – 83.

Zhang, Y. & Mi, Y. (2010). Another Look at the Language Difficulties of International Students. *Journal of Studies in International Education*, 14 (4), 371 – 388. DOI: 10.1177/1028315309336031

APPENDICES

Appendix A – Student Participant Information Sheet

<div style="border:1px solid; padding:8px; text-align:center;">

Information Sheet

</div>

Research Project Title: Testing the untestable or defending the indefensible?: an into EAP assessment usefulness

You are being invited to take part in a research project about the usefulness of academic English language assessments at [MUC]. Before you decide to take part, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this. My contact email is j.clifford2@[muc].ac.uk

The purpose of the project is to investigate the quality and usefulness of the English language assessments in use at [MUC]. This is being done to establish whether students are being equipped with the skills they need to cope with undergraduate and postgraduate level study at the University. The project will take approximately two years to complete.

You have been chosen to participate as you have previously studied on an English language programme at the University. You are being asked to participate in a focus group with other students who have studied on English language programmes; there will be a maximum of six people in your focus group.

It is your choice to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason to withdraw from the study.

If you choose to take part, you will be asked to participate in one focus group discussion. You will be involved in the research until it is completed. You will need to participate once in a focus group of no more than one and a half hours in length. You will be given a copy of the focus group schedule to read through before you take part in the focus group and you will have the chance to ask questions about this schedule before the focus group discussion begins. During the focus group, you will be asked to take part in two small activities and then to discuss these and respond to and discuss the questions on the schedule, plus any follow-up questions which the researcher deems necessary during the discussion. To allow for transcription of your responses, the focus group will be video-recorded and the results of your activities will be photographed. Following the focus group, you will be asked to check the transcription of your contribution, and may be asked for follow-up information should it be required. The video recording and photographs and transcriptions of your responses, along with any comments which you may add if and when asked for further clarification post focus group will be stored securely on a computer which is password-protected and only accessed by the researcher.

The recordings of your activities made during this research will be used only for analysis and for illustration in conference presentations and lectures. No other use will be made of them without

your written permission, and no one outside the project will be allowed access to the original recordings.

Since the research will be taking place on [MUC] premises at a time to suit all participants, no travel expenses will be available to you.

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will enhance the work carried out by the English language department in supporting both new and current students prior to and during their studies on the various programmes at [MUC].

Should the research project stop earlier than expected, you will be contacted by the researcher and the reasons for this will be explained to you.

If you wish to make a complaint, you should contact the project supervisor, Dr. Mark Payne in the first instance. However, should you feel that your complaint has not been handled to your satisfaction, you can contact the University's 'Registrar and Secretary'.

All the information that is collected about you during the course of the research will be kept strictly confidential. You will not be identifiable in any reports or publications.

The results of the research will be included in a thesis to be submitted in partial fulfilment of the degree of Doctor of Education at the University of Sheffield. A copy of the completed study will be available via the researcher should you require one. The focus group results will be presented in aggregated form, so it will not be possible to identify you personally from what you say.

This project has been ethically approved via the University of Sheffield Education Department's ethics review procedure. The University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review Procedure across the University.

For further information, please contact Dr. Mark Payne, Lecturer in Language Education
Department of Educational Studies, University of Sheffield, The Education Building,
388 Glossop Road, Sheffield. S10 2JA UK

Email: mark.payne@sheffield.ac.uk
Tel: 0114 2228170

You will be given a copy of this information sheet and a signed consent form.

Many thanks for agreeing to take part in this research project.

Appendix B Subject-Specialist Staff Participant Information Sheet

Research Project Title: Testing the untestable or defending the indefensible?: An investigation into EAP assessment usefulness.

You are being invited to take part in a research project about the usefulness of academic English language assessments at [MUC]. Before you decide to take part, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this. My contact email is j.clifford2@[muc].ac.uk

The purpose of the project is to investigate the quality and usefulness of the English language assessments in use at [MUC]. This is being done to establish whether students are being equipped with the skills they need to cope with undergraduate and postgraduate level study at the University. The project will take approximately two years to complete.

You have been chosen to participate as you are a member of staff who has considerable experience in teaching international students at the University. You are being asked to participate in a focus group with other members of staff who have similar experience to you; there will be a maximum of six people in your focus group.

It is your choice to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason to withdraw from the study.

If you choose to take part, you will be asked to participate in one focus group discussion. You will be involved in the research until it is completed. You will need to participate once in a focus group of no more than one and a half hours in length. You will be given a copy of the focus group schedule to read through before you take part in the focus group and you will have the chance to ask questions about this schedule before the focus group discussion begins. During the focus group, you will be asked to take part in two small activities and then to discuss these and respond to and discuss the questions on the schedule, plus any follow-up questions which the researcher deems necessary during the discussion. To allow for transcription of your responses, the focus group will be video-recorded and the results of your activities will be photographed. Following the focus group, you will be asked to check the transcription of your contribution, and may be asked for follow-up information should it be required. The video recording and photographs and transcriptions of your responses, along with any comments which you may add if and when asked for further clarification post focus group will be stored securely on a computer which is password-protected and only accessed by the researcher.

The recordings of your activities made during this research will be used only for analysis and for illustration in conference presentations and lectures. No other use will be made of them without

your written permission, and no one outside the project will be allowed access to the original recordings.

Since the research will be taking place on [MUC] premises at a time to suit all participants, no travel expenses will be available to you.

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will enhance the work carried out by the English language department in supporting both new and current students prior to and during their studies on the various programmes at [MUC].
Should the research project stop earlier than expected, you will be contacted by the researcher and the reasons for this will be explained to you.

If you wish to make a complaint, you should contact the project supervisor, Dr. Mark Payne in the first instance. However, should you feel that your complaint has not been handled to your satisfaction, you can contact the University's 'Registrar and Secretary'.

All the information that is collected about you during the course of the research will be kept strictly confidential. You will not be identifiable in any reports or publications.

The results of the research will be included in a thesis to be submitted in partial fulfilment of the degree of Doctor of Education at the University of Sheffield. A copy of the completed study will be available via the researcher should you require one. The focus group results will be presented in aggregated form, so it will not be possible to identify you personally from what you say.

This project has been ethically approved via the University of Sheffield Education Department's ethics review procedure. The University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review Procedure across the University.

For further information, please contact Dr. Mark Payne, Lecturer in Language Education
Department of Educational Studies, University of Sheffield, The Education Building,
388 Glossop Road, Sheffield. S10 2JA UK

Email: mark.payne@sheffield.ac.uk
Tel: 0114 2228170

You will be given a copy of this information sheet and a signed consent form.

Many thanks for agreeing to take part in this research project.

Appendix C Permanent English Teaching Staff Participant Information Sheet

Research Project Title: Testing the untestable or defending the indefensible?: An investigation into EAP assessment usefulness.

You are being invited to take part in a research project about the usefulness of academic English language assessments at [MUC]. Before you decide to take part, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this. My contact email is j.clifford2@[muc].ac.uk

The purpose of the project is to investigate the quality and usefulness of the English language assessments in use at [MUC]. This is being done to establish whether students are being equipped with the skills they need to cope with undergraduate and postgraduate level study at the University. The project will take approximately two years to complete.

You have been chosen to participate as you are a member of the English teaching team at the University. You are being asked to participate in a focus group with other members of the English teaching team, and the group will be made up of no more than six people in total.

It is your choice to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason to withdraw from the study.

If you choose to take part, you will be asked to participate in one focus group discussion. You will be involved in the research until it is completed. You will need to participate once in a focus group of no more than one and a half hours in length. You will be given a copy of the focus group schedule to read through before you take part in the focus group and you will have the chance to ask questions about this schedule before the focus group discussion begins. During the focus group, you will be asked to take part in two small activities and then to discuss these and respond to and discuss the questions on the schedule, plus any follow-up questions which the researcher deems necessary during the discussion. To allow for transcription of your responses, the focus group will be video-recorded and the results of your activities will be photographed. Following the focus group, you will be asked to check the transcription of your contribution, and may be asked for follow-up information should it be required. The video recording and photographs and transcriptions of your responses, along with any comments which you may add if and when asked for further clarification post focus group will be stored securely on a computer which is password-protected and only accessed by the researcher.

The recordings of your activities made during this research will be used only for analysis and for illustration in conference presentations and lectures. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings.

Since the research will be taking place on [MUC] premises at a time to suit all participants, no travel expenses will be available to you.

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will enhance the work carried out by the English language department in supporting both new and current students prior to and during their studies on the various programmes at [MUC].
Should the research project stop earlier than expected, you will be contacted by the researcher and the reasons for this will be explained to you.

If you wish to make a complaint, you should contact the project supervisor, Dr. Mark Payne in the first instance. However, should you feel that your complaint has not been handled to your satisfaction, you can contact the University's 'Registrar and Secretary'.

All the information that is collected about you during the course of the research will be kept strictly confidential. You will not be identifiable in any reports or publications.

The results of the research will be included in a thesis to be submitted in partial fulfilment of the degree of Doctor of Education at the University of Sheffield. A copy of the completed study will be available via the researcher should you require one. The focus group results will be presented in aggregated form, so it will not be possible to identify you personally from what you say.

This project has been ethically approved via the University of Sheffield Education Department's ethics review procedure. The University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review Procedure across the University.

For further information, please contact Dr. Mark Payne, Lecturer in Language Education
Department of Educational Studies, University of Sheffield, The Education Building,
388 Glossop Road, Sheffield. S10 2JA UK

Email: mark.payne@sheffield.ac.uk
Tel: 0114 2228170

You will be given a copy of this information sheet and a signed consent form.

Many thanks for agreeing to take part in this research project.

Appendix D – Agency English Teaching Staff Participant Information Sheet

Research Project Title: Testing the untestable or defending the indefensible?: An investigation into EAP assessment usefulness.

You are being invited to take part in a research project about the usefulness of academic English language assessments at [MUC]. Before you decide to take part, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.   My contact email is j.clifford2@[muc].ac.uk

The purpose of the project is to investigate the quality and usefulness of the English language assessments in use at [MUC]. This is being done to establish whether students are being equipped with the skills they need to cope with undergraduate and postgraduate level study at the University. The project will take approximately two years to complete.

You have been chosen to participate as you are a member of the English teaching team at the University. You are being asked to participate in a focus group with three other members of the teaching team who work for the University but who are not employed on a permanent contract.

It is your choice to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason to withdraw from the study. Your decision to withdraw from this study will not impact in any way on your professional relationship with the University.

If you choose to take part, you will be asked to participate in one focus group discussion. You will be involved in the research until it is completed. You will need to participate once in a focus group of no more than one and a half hours in length. You will be given a copy of the focus group schedule to read through before you take part in the focus group and you will have the chance to ask questions about this schedule before the focus group discussion begins. During the focus group, you will be asked to take part in two small activities and then to discuss these and respond to and discuss the questions on the schedule, plus any follow-up questions which the researcher deems necessary during the discussion. To allow for transcription of your responses, the focus group will be video-recorded and the results of your activities will be photographed. Following the focus group, you will be asked to check the transcription of your contribution, and may be asked for follow-up information should it be required. The video recording and photographs and transcriptions of your responses, along with any comments which you may add if and when asked for further clarification post focus group will be stored securely on a computer which is password-protected and only accessed by the researcher.

The recordings of your activities made during this research will be used only for analysis and for illustration in conference presentations and lectures. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings.

Since the research will be taking place on [MUC] premises at a time to suit all participants, no travel expenses will be available to you.

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will enhance the work carried out by the English language department in supporting both new and current students prior to and during their studies on the various programmes at [MUC]. Please be assured that this research is in no way an evaluation or assessment of the work you do at the University; it is being carried out for an independent research project.

Should the research project stop earlier than expected, you will be contacted by the researcher and the reasons for this will be explained to you.

If you wish to make a complaint, you should contact the project supervisor, Dr. Mark Payne in the first instance. However, should you feel that your complaint has not been handled to your satisfaction, you can contact the University's 'Registrar and Secretary'.

All the information that is collected about you during the course of the research will be kept strictly confidential. You will not be identifiable in any reports or publications.

The results of the research will be included in a thesis to be submitted in partial fulfilment of the degree of Doctor of Education at the University of Sheffield. A copy of the completed study will be available via the researcher should you require one. The focus group results will be presented in aggregated form, so it will not be possible to identify you personally from what you say.

This project has been ethically approved via the University of Sheffield Education Department's ethics review procedure. The University's Research Ethics Committee monitors the application and delivery of the University's Ethics Review Procedure across the University.

For further information, please contact Dr. Mark Payne, Lecturer in Language Education
Department of Educational Studies, University of Sheffield, The Education Building,
388 Glossop Road, Sheffield. S10 2JA UK
Email: mark.payne@sheffield.ac.uk
Tel: 0114 2228170

You will be given a copy of this information sheet and a signed consent form.

Many thanks for agreeing to take part in this research project.

Appendix E – Student Focus Group Schedule

Introduction:
- Thank you for agreeing to take part in this discussion today.
- Please remember that this is my personal research project, and I am not here to test your English, so don't worry about making mistakes as you speak, don't worry about spelling, and do ask for explanation or clarification as we go along if there is anything you don't understand.
- Before we begin, does anyone have any questions?
- There are two parts to this focus group discussion.
- First of all, I'd like you to write some things down on these sticky notes and then we'll talk about what you have written.
- After that, we will have a discussion about the English language assessments in use here at [MUC].
- I will be making notes to remind me of things as we go along and to help me with my transcription

Before we begin part one, please can you introduce yourself, tell us which programme you are on now (or which programme you did last) and tell us which English language course(s) you studied here before you started your main course
ASK QUESTIONS:
- Name
- What would you like me to call you?
- English prog
- Course now?

Part one
- Write on these sticky notes the English language and academic skills and activities which you think have been important on your course so far. Please put only one skill or activity on each sticky note. TIMER FOR 10 MINS MAX

- Discussing with the whole group, please could you organise your notes into categories on this large sheet of paper? Write the categories on the paper and then put the sticky notes into the appropriate category. It is up to you how you categorise the skills.

Questions about this:
- How did you choose these categories?
- Why have you chosen to categorise these skills like this?
- As a group, can you discuss which (if any) category is most important, and which (if any) skills are most important within those

Part two
[MUC] English language assessments.
- What assessments were used at the end of your English language course?
- How were you assessed (essay, in-class test etc)?
- How did the assessment activities compare to the activities that you have included on this sheet of paper?
- Do you think the assessments used on the English programme were useful?
- Do you think that we tested the right skills in the right way?
- Can you think of an alternative way to assess the academic English skills needed to cope at HE level?
- If you could change the assessments, would you? If so, how? If not, why?

- Do you have any further comments or advice for the English team about the English language assessments?
- Is there anything else you'd like to add?

Thank you very much for your help and co-operation today. Please remember that if you have any questions, just email me.

Appendix F – Subject Specialist Staff Focus Group Schedule

Introduction:
- Thank you for agreeing to take part in this discussion today.
- Please remember that this is my personal research project
- Before we begin, does anyone have any questions?
- There are two parts to this focus group discussion.
- First of all, I'd like you to write some things down on these sticky notes and then we'll talk about what you have written.
- After that, we will have a discussion about the English language assessments in use here at [MUC], which you should have had a chance to look through. There are copies here if anyone needs them as a reminder.

Before we begin part one, please can you introduce yourself, tell us which programmes you teach on now and the main subjects you teach.

Part one
- The first thing I'd like you to do is write on these sticky notes the English language and academic skills and activities which you think are important for students to be able to cope on your courses.
- Please put only one skill or activity on each sticky note.

Now you have these, discussing with the whole group, please could you organise your notes into categories on this large sheet of paper? Write the categories on the paper and then put the sticky notes into the appropriate category. It is up to you how you categorise the skills.

Questions about this:
- How did you choose these categories?
- Why have you chosen to categorise these skills like this?
- Can you explain which of these categories or skills are the most important? Why? Is there a hierarchy?

Part two
[MUC] English language assessments.
The English language assessments and grading criteria currently used at the end of the programmes have already been sent to you.
- Based on what you have seen, do you think the assessments and criteria used are a useful way to assess the skills which you have identified as the most important?
- Do you think that we tested the right skills in the right way?
- Can you think of an alternative way to assess the academic English skills needed to cope at HE level?
- If you could change the assessments, would you? If so, how? If not, why?
- Do you have any further comments or advice for the English team about the English language assessments?
- Does anyone have any further comment?

Thank you very much for your help and co-operation today. Please remember that if you would like to withdraw from the study, or if you have any questions, just email me on the address that is on the information sheet.

Appendix G – English Teaching Staff Focus Group Schedule

Introduction:
- Thank you for agreeing to take part in this discussion today.
- Please remember that this is my personal research project, and will not impact on your work at [MUC]
- Before we begin, does anyone have any questions?
- There are two parts to this focus group discussion.
- First of all, I'd like you to write some things down on these sticky notes and then we'll talk about what you have written.
- After that, we will have a discussion about the English language assessments in use here at [MUC].

Before we begin, please can you introduce yourselves and say which English programmes you have taught on?

Part one
- The first thing I'd like you to do is write on these sticky notes the English language and academic skills and activities which you think are important for students to be able to cope on a HE programme.
- Please put only one skill or activity on each sticky note.

Now you have these, discussing with the group, please could you organise your notes into categories on this large sheet of paper? Write the categories on the paper and then put the sticky notes into the appropriate category. It is up to you how you categorise the skills.

Questions about this:
- How did you choose these categories?
- Why have you chosen to categorise these skills like this?
- Can you explain which of these skills are the most important in each category? Why?
- Do you think any of the categories are more important than others? Are there any skills which you think are more important? Is there a hierarchy?

Part two
- [MUC] English language assessments.
- What assessments have you used at the end of the English language courses you have taught on?
- How were the students assessed (essay, in-class test etc)?
- How did the assessment activities compare to the activities that you have included on the sheet of paper?
- Do you think the assessments used on the English programmes are useful?
- Do you think that we tested the right skills in the right way?
- Can you think of an alternative way to assess the academic English skills needed to cope at HE level?
- If you could change the assessments, would you? If so, how? If not, why?
- Do you have any further comments about the English language assessments?

Thank you very much for your help and co-operation today. Please remember that if you would like to withdraw from the study, or if you have any questions, just email me on the address that is on the information sheet.

Appendix H Data Collection: Sample Photographs from Focus Group Activities

English teaching Staff

Subject Specialist Group 1   Subject Specialist Group 2

Student Group 2

Appendix I – Primary Data Thematic Planning Example

Text highlighted in yellow was from the sticky notes****

Theme	English team
Motivation	
Validity	The validity of some of the tasks being used was questioned. For example, whether or not a student should be expected to make a summary of their notes in a listening exam was queried; should students be required to write a summary of an article in a reading exam. See LM quote below. She makes a good point. This is certainly something that will need exploring further. Further issues with validity came up when the team was discussing the time-scale of the writing assessments. Both SD and LM discussed this point, querying whether five weeks was sufficient for producing a 1000-word essay based on secondary research from a set reading pack. As well as discussing the validity of the assessments, considering the tasks, the English team
Note-taking	<mark>listening and note-taking; using secondary sources; presenting references accurately; listening for main idea/detail; identifying main ideas in text; note-taking; the ability to take notes for a particular topic from extended reading materials; skim, scan, detail; effective note-taking; referencing – recording details accurately</mark>. The ability to take notes crosses over between the reading/writing/listening. There is note-taking in lectures which is linked to the ability to deal with the spoken language. There's then the ability to take notes from written sources which will help support writing or speaking tasks. There's then the ability to actually transfer these notes into these written/spoken tasks. The listening note-taking involves being able to identify the main points within a lecture (BW 223 and SD 231-3). For this group, the note-taking also linked to reading comprehension since a student needs to have a certain level of comprehension to be able to take notes in the first place (SD 605) CL 678-682 – the note-taking also includes an element of being able to sift through a large amount of material in order to find the most useful or relevant information. This could be oral or written information. This note-taking is repeatedly discussed in this FG. In describing the listening tasks, SD and CL 965 – 969 explain that the tasks in the assessments are based on lecture-listening (albeit quite short with a clear structure that might not be the case in 'real life'). There is also a requirement to write a summary of the lecture. There was some disagreement in this FG as to whether this summary skill was actually necessary once the students move on since as BW 973-4 asks 'who needs summaries? When do they ever actually have to do a summary of the lectures. However, the discussion continues, with SD emphasising that the ability to summarise the lectures is another way of testing listening comprehension, but which doesn't mean the students can get away with simply identifying the odd key word and guessing an answer. Although perhaps lacking in authenticity, this type of task is a valid assessment of listening comprehension. LM (989-992), however, does question the validity of this type of summary task, asking whether it is testing just the listening skills. She says 'once you then bring a summary into it as well then are you testing something else? Are you actually starting to test language skills and structuring skills again? Are you testing writing rather than listening?' Later she says (LM 1033) 'that's testing being able to write from your notes' and (1035-6) 'if you've made the notes in the first place, that shows me whether you've understood it.
Paraphrasing	<mark>an ability to synthesise numerous sources into a coherent piece of writing (paraphrasing); a wide range of synonyms for use in academic work (paraphrasing); a range of strategies for referring to authors inside the students' written work, to refer to research done (paraphrasing); referencing; the ability to write reasonably coherently with regard to grammar (like wot I haven't!); a wideish range of academic</mark>

294

	<mark>style sentence structures; paraphrasing/summary; an awareness of word forms and ability to manipulate them (paraphrasing); paraphrasing skills; ability to find references for their work from whatever source; ability to paraphrase (v. difficult) and reference; ability to assimilate what they have read into their writing; paraphrasing skills.</mark> This could also be taken to mean effective use of secondary sources. BW 78 discusses 'paraphrasing and assimilating what they've learnt into their work'. LM 84 explains paraphrasing as also being a way to demonstrate understanding of the materials that students have been reading. The paraphrasing/using secondary sources prompted a lot of discussion in this group throughout the FG session. CL 100-1 stated that the use of secondary sources not only demonstrates understanding, it also demonstrates 'the ability to select information to communicate an idea which is a communicative skill rather than a purely more understanding based skill'. It seemed throughout this FG that paraphrasing or at least effective use of secondary sources was a particularly important skill for students. Hence, giving students an 'authentic' assessment experience in which they are required to read, assimilate and make effective use of secondary sources to support opinions expressed in essays and reports. There is a definite difference between the use of references in work and the mechanics of presenting references accurately. It was agreed that even in a short language course (5 weeks, as is the SEP), it would be possible to assess the ability to effectively integrate secondary sources, even if it is just one or two. These references can, as they currently are, be provided by the team as set reading to make the life of students easier.
Language skills	<mark>cohesion, linkers; AWL items; conditionals; complex noun phrases; knowledge of subject specific technical vocabulary; knowledge of academic vocabulary – awareness of some of the subtleties; extended noun-clause recognition; problem-solution-evaluation; certain subject spec vocabulary; academic register; agreeing disagreeing opinion; compare contrast; identifying lecture lang; sentence syntax; relative clauses; understanding parts of speech (and their purpose); complex noun phrases; relative clauses; describing data accurately Cf <u>trends</u>; hedging; hedging language</mark> Key language for academic use arising from this discussion included hedging language (LM 32). Other terms arising from this discussion include: • (CL76) grammar knowledge • (CL76) knowledge of vocabulary • (SD99) using synonyms • (SD99) restructuring language • Language 'skills' – L,R,W,S • (LM171) relative clauses • (CL277) 'chunks' of language (lexis and grammar) • (SD280) register and style • (SD287) modals and grammatical writing • Fluency – both spoken and written • (SD637) reporting verbs There was some serious debate as to whether to split the categories into the four key skills areas, but that was decided against owing the interconnected nature of the language for academic purposes. A further area discussed under this heading was describing data, since as LM294 says 'it's sort of grammatical'. This is supported by SD295 who agrees that in describing, for example, trends, 'they've got to command tenses' It was agreed that the language side of EAP includes 'grammar and academic style' (LM324) and 'academic register' (CL370). Language skills, as well as covering the productive side, such as being able to write or speak, also include the receptive, comprehension skills. And these are then combined when the students read to write or read to speak.

	By way of a summary for this section, CL663-6 says 'so we've got language skills which covers a whole bunch of grammar stuff. It also covers academic language and subject-specific language. We've got structural language such as hedging, such as showing agreement and disagreement. We've got formality, so academic register' SD 667 'Academic word list' 669 'complex noun phrases' CLl670-1 'which is the grammatical features. … so there is specific grammatical features' A concern raised by the team which has led to an idea from me is whether or not there is sufficient focus on the grammar and vocabulary? Have the assts now moved so far towards the academic skills things that we are now missing some key elements from the assessments? Perhaps we do need to not necessarily re-write the assessments, but re-consider the grading criteria which are being used? Do we need more focus in these on the grading criteria? Would this be a way of enhancing the validity of the assessments in this regard? CL 1225 -6 explains 'we had a whole bunch of grammar skills and vocab skills … what's happening with the focus on those particular skills in our assessments?' LM1363 also asks 'Should we be just language?' CL1364-70 responds to this stating 'Yeah, but then what you're saying is that we're looking at what they need to do to be on that course and we're saying that they need the language skills ultimately is the priority to be on that course, and our purpose is to test the language. So again, we're going to define what we're testing by where they're going to say well, it would be nice if we could also introduce the life skills, the criticality , the other stuff, but primarily we're doing language and we're going to define progression onto that course by the language.' LM1483-6 'dependent on length of course we need to be mindful of how deeply we can go into the academic skills side because there's a danger that by focussing too much on academic skills we miss the nuts and bolts of ensuring that students understand key vocabulary'. She goes on to highlight the issue that we might 'overlook the immediate language needs' (LM1492-3). There was definitely an issue with the lack of grammar that we teach them. BUT if the purpose of the course is to prepare students to go on to UKHE, then there needs to be the language element as well as the academic skills side of things. A skill that actually came up during the discussion of life skills area which could actually be considered a language skill is the ability to write concisely, i.e. to present ideas in bullet point form (see SS1 discussion too as this is discussed there).
Speaking	<mark>presentation skills; clarification/comprehension; referring to research in debate; turn taking; fluency;</mark> CL360 'the question is, what's speaking?' this is a very important question. HOW do you categorise speaking for academic purposes? I would say that the speaking involve academic speaking tasks, but at the same time, there are other needs for speaking skills that are not academic but which are encountered in an academic context. This is discussed by the SS FGs. Key academic speaking skills include seminar and presentation skills, be it individual or in a group, and debating. The speaking assessments that have been used have been varied. These assessments have ranged from IELTS style interviews, debates, seminars, individual presentations, group presentations and digital portfolios. CL671 by way of summarising this section 'We've got the speaking skills where we've got referring to research when you're talking' CL673-4 'Clarification and comprehension which also links in with things like the turn-taking skills. There's presentation skills, and we've got fluency. We probably could've added in a few more thing there'. BW675-6 'Yeah coz we started talking about speaking for seminars and stuff like that so there's probably a bit more isn't there?'
Critical stuff	<mark>Critical thinking/reading skills; understanding feedback; effective reading skills; critical thinking; ability to read critically; recognition of the value and dangers of non-academic sources; awareness of a variety of academic styles of writing *purpose *structure; proofreading/redrafting; ability to reflect on their (to some degree) work; self-correction/proofreading</mark> The 'critical stuff' category was an interesting one. It was clearly considered an important area by the team, yet the title is as un-academic as you can get!

	However, the term was agreed on as being the best way to categorise this area of approaching work critically and having an element of criticality to the tasks being set. This critical stuff category was returned to several times during the first part of the discussion in the categorisation activity, starting with SD89 who said ' if you just directly quote something it doesn't mean anything does it?' adding that when incorporating references, students are 'maybe using that as a way of commenting on an idea as well rather than just here's a quote' (SD103-4). CL also looks at reading and using the work of others in a similar way stating that 'showing support of arguments' is important, and how students should be using their reading (CL190-1). LM152-3 also brings in the critical stuff quite early explaining that reading critically is important for being an effective student. Further discussion on this included: CL248 'Assimilating comes after you've done the research' BW249 'Coz you need to know what you're assimilating, don't you. You need to know what's worth it.' A further area that comes under this category is the ability to reflect on your work (SD455) and it was emphatically agreed by the group that this was definitely a critical skill. SD (461 and 463) asks a helpful question about the critical side of things 'And would proofreading and redrafting be critical as well? Coz as you're doing the draft you're picking it apart you're saying ok, that works, that doesn't'. A further area of critical stuff as defined by CL765-7 who says 'Recognising the different types of things that you're reading and things that you're thinking about, well, what are we doing with it? We've got something that's non-academic, but how can we evaluate it? How can we be aware of how to use it?' SD868-71 and 873-6 says, quite interestingly 'When I probably first came here I would've thought in those lines of writing, speaking, listening, but the longer I've been engaged with research myself and supervising dissertations and reading the dissertations I've sort of drifted more towards thinking well the critical component is a very important aspect of, and usually lacking. I think it just as kind of your career develops as well and you're teaching develops and the more you see of the work that, or lack of it, that goes on at the college, your viewpoint shifts and you zero in on different areas of the academic skills sets.
Life skills/ study skills	autonomous learning; time management; time management; IT skills input for research (and generally I suppose); ability to study and balance social life/work; IT skills; designing effective visual aids; ability to budget so they don't run out of money; working independently – autonomy; presenting data in tables, charts, etc; awareness of diff teaching style/expectations Additional skills that came up during the categorisation section and the discussion were: • Beginning to be aware of different teaching styles and expectations • Adapting to or awareness of a different academic environment • Effective use of presentation software to structure and support a presentation • Effective use of IT skills to present essays, to analyse data and create charts and tables to effectively present these data There is a cross-over between the IT skills and the research skills since without the IT skills then the students can't do the research. They will struggle to be able to access the materials that they need to do their work. Even access to the VLE is vital since, as SD explains, lecturers post links and articles on there. In summarising this category, BW explains how it was hard to decide exactly what to put here since there is so much cross-over between the categories they used: 790-1 'and they you have the IT skills and designing effective visual aids which came here finally after much kicking and screaming'. A skill that I have added to this is how lectures are used to inform studying; how they are used to help the student develop their ideas and own opinions in the first place. This is tested in some of the listening assessments in which the students are asked to summarise the content of the lectures they are listening to in the exams.

	This summary skill, and ability to transfer what they are hearing into re-formulated sentences at a later date is both a language and a study skill. CL984 'I think that's encouraging them to select the important information from the lecture'. SD 985 'and it also demonstrates, we were talking about, well what's the point of paraphrasing is it a technical exercise is it a comprehension exercise, what is it?' [pick this point up somewhere else – is needed…] Encouraging good study skills as CL 1001-2 says is good practice, no matter how bad the lectures are at the institution the students are going into 'but we're encouraging good life skills and I think what we see in lectures isn't necessarily a reflection of good… Of the best practice in lectures' Further examples of life skills from the assignments students have to do is when they develop their further technical competences when they are required to record the digital portfolios. Developing autonomy was another area that the EDC believe should be assessed (somehow) – or perhaps this is one of the key academic competences that can be developed through the assessments rather than directly assessed. Autonomy is something really that needs to be fostered rather than assessed directly, as do a lot of these life skills.
Structure	<mark>discourse structure; ability to organise their texts logically and coherently; selecting and putting together data, graphs, etc.; structuring assignments; paragraph structure; structuring presentation; writing genre conventions; the writing process: *planning *producing *reviewing; structuring paragraphs (different types)</mark> This is perhaps one of the academic competences rather than a skill to be assessed. Structure links to both spoken and written language Is this included in our assessments? Do we assess for structure (for either language skill)? Structure links to writing genre conventions (SD 738) In language testing, we always 'wanna see a well-structured essay' (LM1427-8) Structure, although a category outlined by the FG was not really discussed in much detail. Perhaps this could be absorbed into another category eventually? Essentially, well-structured writing tasks and presentations are key to coping with HE level study, however the structure is perhaps one of the life skills rather than being a separate category?
Understand-ing assignments	<mark>understanding assessment criteria; decoding assignment remits; understanding different genres of written and spoken texts; ability to break down assignments into their constituent parts; understanding assignments; understanding academic terminology e.g. research/methodology/evaluate</mark> This links to understanding key assignment task words such as discuss and evaluate as well as understanding academic terminology. By including these types of words in the EAP tasks that students are assessed with. The ability to decode the assignment remit can be addressed by giving the EAP assessments a level of presentation authenticity (or should that be face validity?) so that they look similar to assts used elsewhere in the institution and therefore give students experience of decoding them (especially since they can be very complicated at the case study HEI). Further authenticity can be given by using similarly structured grading criteria and by using that terminology which has been identified as key by the test users. This section also linked in understanding feedback – this was discussed by other FGs too. By offering students feedback presented in a similar way to that which they will receive on their main courses, the grading criteria can be considered valid. However, it is only by including some element of reflection into the assessment tasks that the grading criteria could be considered useful. This is touched on by the other FGs too – the need for reflection on feedback. A more formative approach to assessment might, then, be better. CL explains (707 and 709) that 'feedback is part of the assessment. If you don't understand the feedback then you're not going to be able to improve for the next assessment'. BW however disagreed, (722-3) 'I kind of thinking of [understanding assessments] as breaking it down when you get it rather than the post assessment'. CL responded (726, 728) 'But I think understanding the feedback is again breaking down what the assessment is about' BW (729-31) 'It's not you breaking it down, it's someone doing it for you though, isn't it? I think

	[understanding assessment] as the things which you are contributing to it, things that you are doing, whereas feedback is kind of external'
	This argument was interesting, and could perhaps lead me to breaking down the category into understanding assessments and then, linking to the SSFGs, an understanding feedback, using feedback or reflective learning section.
	There was a lot of discussion about academic terminology under this heading, although I'm not entirely certain that this is the best place for it, since words such as primary and secondary research, methodology, quantitative, qualitative are perhaps very specific academic terminology which should have their own category?
	When asked whether the categories might have a hierarchy, LM (899-900) states 'I think the understanding assessment would actually by the highest because actually I think that if you really did understand what was required then you might realise that criticality is required' (904-8) 'But I think if you don't understand what's required, and I would argue that more so than the students driving what we're doing it's actually the assessments, it's our understanding of what is required in the assessments that has driven a lot of the change that's taken place in our department.
Dig out/ research skills	<mark>using databases/library; research skills; selecting key words for research; using VLE; understanding methodology; an ability to scan through extended reading material for key words related to a topic of interest; research skills</mark>
	SD 239-40 'I think for me research skills involves being able to use databases and the library and being able to pick out resources'. This being able to choose appropriate resources could come under the critical stuff heading. This being able to find key information is a vital study skill as well – could it be absorbed into that area too?
	LM 267-69 'Selecting key words for research, that's like actually deciding what your key words will be to do your search. I don't know about this though using secondary sources coz I think that might come into paraphrasing and referencing.'
	In summarising this category, SD 683-85 'Over this side we've got research skills, we kind of debated that a bit, but I think we defined that as actually digging out information, so that would be using databases, library, selecting key words for research, using VLEs [study skill??], but also understanding concepts within methodology' CL adds (686-7 & 698-91) 'so this is the stuff that in a methodology in a dissertation it's what you'd talk about there. That's what we're defining as research. So, if you're talking about research skills, it's almost like we said, this is what we'd do, this is the areas we'd look at in a methodology of a dissertation and that's what we're defining as the research skills separate from the language skills of note-taking or paraphrasing' – this is further indication of the interconnectedness of the whole thing.
	Further examples of what could be included in this section: sampling;
Mechanics	This is not a separate category from the FG, however it does link to the referencing/use of secondary sources above. It is more of an attention to detail type skill. This could be defined as the technical side of compiling a document or a .ppt, or reference list and the more mechanical side of accurately presenting in-text citations.
Interconnec-tedness	Throughout this discussion it became clear that it is just impossible to separate the skills and the language in an academic context. Concrete examples include:
	• The lack of clarity with regards to what is language skills and what is life skills (SD 609 – SD 612).
	• Whether note-taking should be a separate skill area or whether it should be considered part of the broader area of language.
Presenting research	This is another new category not defined by the FG. However, I think it emerges during the discussion that how to present research, how to effectively write about it, is a key skill and the language needed to do this effectively needs to be assessed in some way since this is what the students will be required to do at some level during their studies.
Understand-ing feedback	This could come in the understanding assessments section too, but might also be a critical skill… CL715-8 'reflection on your own work is looking at what you did and evaluating what you did, whereas feedback is looking at what the lecturer's looking and going 'Ah, they were looking for X, I put in Y'.

	SD751-3 explains that this understanding feedback area could be in the critical area since 'I don't think that it should be just a one-way process. I think you get feedback from the lecturer and In some, at some level you engage with it.' SD756 'And you kind of question well 'Have I been doing things right?'' which is, in a way SD759-60 'encouraging a new line of thinking'
Current assessments	Current speaking assessments in use include the digital portfolio. SD1098-1105 'it's an idea I kind of stole from someone who talked about it at IATEFL. Basically, one week you'll focus on a particular topic and you would get certain language, agreeing, disagreeing, opinion and they you'll give them a task to do with a partner and record it … and they submit three over the semester and those are worked out as 30% of the overall mark and then the in-class seminar. I think it's so that there's ongoing feedback into their seminar. CL agrees (1107) that the students 'get a lot of formative feedback from it' and SD 1108-9 adds that 'it caters for the quieter students, students that perhaps need to build their confidence in terms of engaging with other students in speaking.' I have questioned the purpose of the assts being used. The multi purpose of the tests is queried – are they EAP proficiency or summative tests? Because they also appear to be, to some extent, both formative in that they offer students comments on their weaknesses and how they can improve them, whilst also being diagnostic in that we are diagnosing student weaknesses in order to decide on progression. The way the critical skills are tested are various. In two of the writing assessments used, students are asked to include an evaluative literature matrix in which they demonstrate not only their understanding of the text, but also their ability to critically evaluate the texts for use in their essay/report. This critical skill implies a level of comprehension on the part of the students since they cannot evaluate the sources if they cannot understand them in the first place. So, my question is, does reading comprehension need to be assessed separately if they students can demonstrate this by the way the sources are used and evaluated in their writing?
The 'How' of testing Authenticity?	Another new category from my reading. It concerns the way the tests are testing the skills and attributes which were deemed necessary to cope by the EDC. Essentially, the team agreed that so long as the tests are testing the right skills in the right way, then they are useful. The validity and reliability, however, are another question to be dealt with in the appropriate spaces above. The authenticity argument has possibly become weaker here since it is by trying to ensure that the students have an authentic assessment experience that we are moving away from testing pure language and grammar. That said, since language is a contextual beast, there needs to be some level of context to the assessments being used. As this is EAP, it is an academic context that is being used, hence the assessments need to be getting the students to use the language in as authentic a way as possible.
GENERAL THOUGHTS NOTED DOWN	Is it the need for students to understand assessments that has led the EDC to develop and design the assessments in the way that we have? And does this, therefore, lead to the assumption that, to some extent as least, our assessments are useful in that they are giving students experience of the assessment type that they will be facing when they move on? Are we giving them an 'authentic' assessment experience in some cases? What does seem clear from this FG, and might arise from the others, is that where English skills end and where study skills begin is almost impossible to separate. This is an issue for the English team in that we are increasingly expected to test students' English levels, however, we are also being told now that we cannot cross the line into the study skills area since that is covered by another department. However, we are acting as an academic preparatory department, so ignoring these key academic skills in our assessments could lead to a validity issue.

Is the curse of EAP the need to be dual-purpose? Do we need to do the two things? Is it an impossible task?

Developing reflective learning skills and assessing other key skills such as autonomy and general study skills appears to fall under the remit.

SD887-888 in talking about the critical issues that the students have, states that our students 'perhaps aren't used to engaging in seminars or being critical. Not saying that the students we get are never critical in any context but perhaps in educational contexts they're not encouraged to be critical'. BW agrees that it is this area that needs the most work for our students since it's the area that's the most underdeveloped. This links to my idea of the fact that we are expected to essentially undo 18+ years of schooling and educational culture in 5 weeks in some cases to have them ready to go on to their next course.

The more exposure to and contact with the wider university community that the department has had, the more we have felt compelled to change our assessments to match what is being done elsewhere. For example, asst remits, generic criteria, writing and researching rather than pure in-class tests and end of course exams. Whether or not these are a more useful approach to assessing language for academic purposes remains unclear at this point, but perhaps with some further reading into the area, then this might become clearer. SD does support this 936-9 'involvement in dissertations, you know, with the use of generic grading sheets and son on we're kind of exposed more and more to what's going on across college and what the needs are across college so you kind of tweak and redirect and recalibrate what you're doing.' This does seem to indicate perhaps greater input from the SS when it comes to developing useful EAP assessments since they know what is needed once the students move on.
This is also commented on by LM 925-8 'and so with us having a better understanding or realisation, epiphany or something about, ok, what is it that they're doing beyond us? …maybe we started to stretch our vision beyond what we were just doing in the English language team and recognising the [further] needs' 930-32 'we have a better understanding [now] of what it is they're going to have to do and with us being engaged in dissertation supervision, we know what the end goal is gonna be, a clearer idea.

Is the purpose of EAP assessment, then, more than just to test an acquired language skill? By asking students to do this variety of assessments, they are not only being assessed in the various language and academic competences, they are also developing life skills – e.g. how to present a piece of academic writing, how to give a presentation, how to use IT skills. So, the purpose of the assessments is threefold, since as well as assessing the students the students are also being given the opportunity to develop various competences. Should this be another section that is actually added to my model at the end of the thesis? As well as the three areas of authenticity, validity and reliability, should 'opportunity to develop further competences' be included? Perhaps rather than having a set of 'graduate attributes' EAP needs to have a set of 'academic competences' which need to be developed by the assessments being used. Perhaps a set of these could be developed by me to be included at a later date. Perhaps I need to put these in my recommendations? So how do I relate this back to the assessment usefulness areas? For this discussion link to: pedagogical tasks and activities (what are they and are they reflected?). These pedagogical tasks/activities are essentially a way of classifying the wider academic competences that are being developed.
A further area that came up for discussion was the idea of pitching assessments at the right level. This could come under validity, perhaps? LM 944-6 testing the right skills for the courses that they're going on to whist also having to balance the issue of multiple courses being served by one in-sessional group – would 'best fit' be something worth considering?

Appendix J - CEFR Writing Descriptors (Adapted from Council of Europe, 2001)

Level	Holistic description	Reports and Essays
C2	Can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points.	Can produce clear, smoothly flowing, complex reports, articles or essays which present a case, or give critical appreciation of proposals or literary works. Can provide an appropriate and effective logical structure which helps the reader to find significant points. Can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues.
C1	Can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion.	Can expand and support points of view at some length with subsidiary points, reasons and relevant examples.
B2	Can write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesising and evaluating information and arguments from a number of sources.	Can write an essay or report which develops an argument systematically with appropriate highlighting of significant points and relevant supporting detail. Can evaluate different ideas or solutions to a problem.
		Can write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options. Can synthesise information and arguments from a number of sources.
B1	Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence.	Can write short, simple essays on topics of interest. Can summarise, report and give his/her opinion about accumulated factual information on familiar routine and non-routine matters within his/her field with some confidence.
		Can write very brief reports to a standard conventionalised format, which pass on routine factual information and state reasons for actions.
A2	Can write a series of simple phrases and sentences linked with simple connectors like 'and', 'but' and 'because'.	No descriptor available
A1	Can write simple isolated phrases and sentences.	No descriptor available

Appendix K – IELTS Writing Descriptors (IELTS, 2016)

Band	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
9	• uses cohesion in such a way that it attracts no attention • skilfully manages paragraphing	• uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	• uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	• sequences information and ideas logically • manages all aspects of cohesion well • uses paragraphing sufficiently and appropriately	• uses a wide range of vocabulary fluently and flexibly to convey precise meanings • skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation • produces rare errors in spelling and/or word formation	• uses a wide range of structures • the majority of sentences are error-free • makes only very occasional errors or inappropriacies
7	• logically organises information and ideas; there is clear progression throughout • uses a range of cohesive devices appropriately although there may be some under-/over-use • presents a clear central topic within each paragraph	• uses a sufficient range of vocabulary to allow some flexibility and precision • uses less common lexical items with some awareness of style and collocation • may produce occasional errors in word choice, spelling and/or word formation	• uses a variety of complex structures • produces frequent error-free sentences • has good control of grammar and punctuation but may make a few errors
6	• arranges information and ideas coherently and there is a clear overall progression • uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical • may not always use referencing clearly or appropriately • uses paragraphing, but not always logically	• uses an adequate range of • vocabulary for the task • attempts to use less common vocabulary but with some inaccuracy • makes some errors in spelling and/or word formation, but they do not impede communication	• uses a mix of simple and complex sentence forms • makes some errors in grammar and punctuation but they rarely reduce communication
5	• presents information with some organisation but there may be a lack of overall progression • makes inadequate, inaccurate or over use of cohesive devices • may be repetitive because of lack of referencing and substitution • may not write in paragraphs, or paragraphing may be inadequate	• uses a limited range of vocabulary, but this is minimally adequate for the task • may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	• uses only a limited range of structures • attempts complex sentences but these tend to be less accurate than simple sentences • may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	• presents information and ideas but these are not arranged coherently and there is no clear progression in the response • uses some basic cohesive devices but these may be inaccurate or repetitive • may not write in paragraphs or their use may be confusing	• uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task • has limited control of word formation and/or spelling; errors may cause strain for the reader	• uses only a very limited range of structures with only rare use of subordinate clauses • some structures are accurate but errors predominate, and punctuation is often faulty
3	• does not organise ideas logically • may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	• uses only a very limited range of words and expressions with very limited control of word formation and/or spelling • errors may severely distort the message	• attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	• has very little control of organisational features	• uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling	• cannot use sentence forms except in memorised phrases
1	• fails to communicate any message	• can only use a few isolated words	• cannot use sentence forms at all

Appendix L – IEP Research Report Grading Criteria

	A	B	C	D	E
Introduction (5%) /100 Weighted mark:	☐ An effective introduction which clearly states the aim ☐ A clear rationale provided for the report ☐ Organisation of the report clearly stated	☐ A good introduction with an aim for the report ☐ An attempt made to provide a rationale for the work ☐ Points to be covered mentioned	☐ An aim is provided, but this may not be clearly stated ☐ Some rationale is provided, but it may not fully justify the report ☐ Some mention made of the points being covered, but this might only be brief	☐ No clear aim stated ☐ Lacks a clear rationale for the study ☐ The organisation of the work is not made clear in the introduction	☐ The work lacks a coherent introduction ☐ There is no attempt made to provide an aim for the study ☐ There is no attempt made to provide a rationale
Literature review matrix and referencing (15%) /100 Weighted mark:	☐ Effective selection and use of support references and quotations conforming to college standards ☐ A clearly organised and well-synthesised literature matrix ☐ Good evaluation of secondary sources	☐ Good selection and use of references and quotations conforming closely to college standards ☐ An organised matrix with some evidence of synthesis ☐ Adequate evaluation of secondary sources	☐ There is a genuine attempt to reference according to college standards ☐ An attempt made to organise and synthesis sources in the matrix ☐ Some evaluation of secondary sources, but this may be quite simplistic in nature	☐ Limited attempt to comply with the college standard ☐ Limited synthesis of sources in the matrix ☐ Only a basic evaluation of some of the sources/not all sources have been evaluated ☐ Limited justification of approaches	☐ secondary sources have not been evaluated ☐ referencing does not comply with college standard ☐ no synthesis of sources ☐ evidence of plagiarism
Methodology (10%) /100 Weighted mark:	☐ Effective justification of approaches based on research methods texts	☐ Good justification of approaches based on research methods texts	☐ Some justification of approaches based on research methods texts	☐ Limited justification of approaches	☐ approaches have not been justified
Findings (15%) /100 Weighted mark:	☐ A clear report of the findings with a logical flow of ideas ☐ Findings section organised with headings to ensure good organisation ☐ Comparative discussion of primary findings with secondary ☐ Primary findings clearly and logically presented using tables and charts where appropriate	☐ Findings are clearly reported ☐ Headings are used, although not always helpful ☐ A good attempt made to compare primary and secondary findings, but this is not always successful ☐ Some use of charts and tables to display primary findings	☐ Some main ideas presented, but these are not always relevant to the task ☐ Headings may be used inappropriately/may be missing ☐ Some comparison of primary and secondary findings, but this is limited ☐ Charts and tables used, but not always effectively	☐ The findings are not clearly or logically presented ☐ Headings have not been used/are confusing ☐ Limited comparison of primary and secondary findings ☐ Charts and tables not used, or used ineffectively	☐ No clear report of findings ☐ Headings not used ☐ No comparison of primary and secondary findings ☐ No use of charts and tables to display primary findings
Conclusions and recommendations (5%) /100 Weighted mark:	☐ Conclusions offer a logical summary of the key findings ☐ Recommendations are clearly based on the conclusions ☐ Recommendations offered are feasible and relate directly to the task	☐ Some logical conclusions presented, but not all of them relate to the findings ☐ Some helpful recommendations, but these might not all link to the conclusions ☐ Some good recommendations, but not all of these might be feasible	☐ An attempt made to conclude the report, but this might not be clearly based on the findings ☐ Some recommendations presented, but these may not be feasible or might not relate to the conclusions	☐ The conclusions do not really relate to the findings ☐ The recommendations are not clearly based on the conclusions ☐ Some of the recommendations are not feasible	☐ No attempt made to conclude the report ☐ No recommendations presented/recommendations might not be related to the topic ☐ Recommendations are not feasible or logical
Lexical resource (20%) /100 Weighted mark:	☐ Uses a wide range of vocabulary with very natural control of lexical features; very few errors present ☐ Uses a wide range of vocabulary fluently and flexibly to convey precise meanings; occasional inaccuracies in word choice and collocation; rare errors in spelling / word formation ☐ Uses sufficient range of vocabulary; may produce occasional errors in word choice, spelling and word formation.	☐ Uses an adequate range of vocabulary for this task ☐ Attempts to use less common vocabulary but with some inaccuracy ☐ Makes some errors in spelling / word formation, but inaccuracies do not impede comprehension	☐ Range of vocabulary is limited, but is adequate for the task ☐ May make noticeable errors in spelling / word formation that impede comprehension	☐ Uses only basic vocabulary – range is very limited; may not be appropriate for task ☐ Has limited control of word formation/ spelling ; errors impede comprehension	☐ unprofessional use of vocabulary and grammar leading to ineffective communication
Grammar range and accuracy (20%) /100 Weighted mark:	☐ Uses a wide range of structures with full flexibility and accuracy ☐ Uses a wide range of structures; the majority of sentences are error-free; makes only very occasional errors ☐ Uses a variety of complex structures; has good control of grammar and punctuation but may make a few errors ☐ Work is edited and proof-read	☐ Uses a mix of simple and complex forms ☐ Makes some errors in grammar and punctuation but they rarely impede communication ☐ Work is edited and proof-read	☐ Uses a limited range of structures ☐ May make frequent errors regarding punctuation and grammar; errors can impede comprehension ☐ Work is edited and proof-read	☐ Limited range of structures ☐ Some structures are accurate but errors predominate, and punctuation is often faulty ☐ Work is inadequately edited and proof-read	☐ unprofessional use of vocabulary and grammar leading to ineffective communication
Organisation and Academic Style (10%) /100 Weighted mark:	☐ Skilfully manages paragraphing; organisation is highly appropriate to genre and linking expressions are highly effective ☐ Sequences information and ideas logically; uses paragraphing sufficiently and appropriately; manages all aspects of cohesion well ☐ Logically organises information and ideas; there is clear progression throughout; uses a range of cohesive devices appropriately although there may be some under- /over- use; presents a clear central topic within each paragraph	☐ Ideas have a logical sequence ☐ Cohesive devices are used effectively, but some errors are present ☐ Uses paragraphing but not always logically	☐ Presents information with some organisation, but there may be a lack of overall progression ☐ Makes inadequate, inaccurate or over-use of cohesive devices ☐ May not write in paragraphs, or paragraphing may be inaccurate (topic sentences may be missing or paragraphs not clearly organised)	☐ Presents information and ideas but these are not arranged coherently; ideas do not flow logically ☐ Very basic use of cohesive devices – these may be inaccurate or repetitive ☐ May not write in paragraphs or their use may be confusing / impede comprehension	☐ disjointed structure ☐ a lack of coherence and cohesion throughout the work ☐ Paragraphing is not effectively used

304

Appendix M – Ethical Approval Letter

The University Of Sheffield.

The

School

Of

Education.

Julia Clifford

Literacy & Language

Head of School
Professor Cathy Nutbrown

School of Education
388 Glossop Road
Sheffield
S10 2JA

8 January 2014

Telephone: +44 (0)114 222 8096
Email: edd@sheffield.ac.uk

Dear Julia

ETHICAL APPROVAL LETTER

Testing the untestable or defending the indefensible?: an investigation of the usefulness of academic English language assessments at a UK HEI

Thank you for submitting your ethics application. I am writing to confirm that your application has now been approved.

We recommend you refer to the reviewers' additional comments (please see attached). You should discuss how you are going to respond to these comments with your supervisor BEFORE you proceed with your research.

This letter is evidence that your application has been approved and should be included as an Appendix in your final submission.

Good luck with your research.

Yours sincerely

Professor Dan Goodley

Chair of the School of Education Ethics Review Panel

cc   Dr Mark Payne

Enc   <mark>Ethical Review Feedback Sheet(s</mark>)