



UNIVERSITY of
STIRLING



Speaking test development in English for Academic Purposes: A pilot study

Author: Rupert Williams

University of Stirling

British Council ELT Master's Dissertation Awards: **Special Commendation**

**MSc in Teaching English to Speakers of Other Languages
and Applied Linguistics**

**Speaking test development in English for Academic
Purposes: A pilot study**

August 2016

Abstract

Global proficiency exams such as IELTS and TOEFL are ubiquitous within English for Academic Purposes (EAP) and are useful yardsticks for assigning grades and determining course entry criteria. However, several commentators (e.g. Alexander et al. 2008, Schmitt & Hamp-Lyons 2015) suggest that while such summative tests are useful, they often do not adequately prepare students for university study. As a result, there is a need for formative tests that may be used with pre-sessional or in-sessional students. What's more, the British Association of Lecturers in English for Academic Purposes (BALEAP) state that being well versed in both test development and use should be an essential skill for EAP teachers (BALEAP 2008). While academic writing is usually given prominence in EAP courses, several studies have highlighted the deficit in speaking ability among international students entering English medium universities (e.g. Jordan 1997, Alexander et al. 2008, de Chazal 2014). Speaking is an important skill for international students: it is frequently part of assessment, especially at postgraduate level, and it is essential for students' integration.

This research follows the evolution of a formative test of speaking for EAP students. It is designed for international postgraduate students entering or studying in English-medium institutions. The test is designed by myself and contains original items and content (such as the use of informal language, using countries as content and including a false assertion by the examiner). The test was designed based on personal experience and the testing literature. It was initially piloted and then distributed to EAP teachers and students throughout the UK via channels such as the BALEAP mailing list. Both students' and teachers' attitudes towards the test were then gained via questionnaires. Questionnaires were used to collect useful data by using both open and closed questions. The data were then analysed using a simple descriptive method.

Findings showed that general attitudes towards the test were positive, such as the belief that the test is valid, reliable and that informal language is relevant for EAP. However, there was some dissonance between attitudes of students and teachers towards certain aspects of the test, such as the attitudes towards the validity of the false assertion and the use of countries as content.

Attitudes gained also highlighted a number of test design issues, such as the need to expand the marking criteria, which could be taken on-board to develop the test further and move towards a viable EAP testing tool.

Acknowledgements

I am grateful to my dissertation supervisor for his keen eye for detail and constructive feedback in putting together this dissertation.

I am also very grateful to mum whose help and support throughout the year has made the completion of my studies much more manageable.

I am also grateful to all my students, colleagues and friends in Galicia, Spain, who welcomed me to their country and gave me the opportunity to develop professionally. I dedicate this dissertation to them, and in particular my small Spanish family.

Glossary

BALEAP British Association of Lecturers in English for Academic Purposes

EAP English for Academic Purposes

ESP English for Specific Purposes

EFL English as a Foreign Language

ESL English as a Second Language

TESOL Teaching English to Speakers of Other Languages

IELTS International English Language Testing System

TOEFL Test of English as a Foreign Language

EMI English medium instruction

ETIC English Teaching Information Centre (*Now part of the British Council*)

EGAP English for General Academic Purposes

ESAP English for Specific Academic Purposes

L1 First language/mother tongue

L2 Second language/medium of communication

List of Figures and Tables

Figure 1. Research timeline showing the order of test development and data gathering.

Figure 2. EAP teachers' attitudes towards the statement: "*The test measures the types of speaking tasks that EAP students will need in their academic studies*".

Figure 3. Bar chart showing students' attitudes towards the statement "*The test measures the types of speaking tasks that are relevant to my academic studies*".

Figure 4. Bar chart showing EAP teacher attitudes towards the statement: "*The content of the test is relevant to EAP students*".

Figure 5. Bar chart showing EAP student attitudes towards the statement: "*The content of the test is relevant to my academic studies*".

Figure 6. EAP teacher attitudes towards the statement: "*Both candidates have the same opportunity to demonstrate their speaking ability*".

Figure 7. EAP student attitudes towards the statement: "*I had the same chance as the other candidate to demonstrate my speaking ability*".

Figure 8. Teacher attitudes towards the statement: "*If the test was mandatory, it would require specific test preparation classes*".

Figure 9. Student attitudes towards the statement: "*If I did the test again, I would like to attend specific test preparation classes*".

Figure 10. Bar chart showing teachers' attitudes towards the statement: "[Informal language] *is appropriate for an EAP test of speaking*".

Figure 11. Bar chart showing students' attitudes towards the statement: "[Informal language] *is relevant to my academic speaking ability*".

Figure 12. Bar chart showing teachers' attitudes towards the statement: "*This false assertion is appropriate for an EAP speaking test*".

Figure 13. Bar chart showing students' attitudes towards the statement: "*The false assertion is helpful in developing my academic speaking abilities*".

Figure 14. Teacher attitudes towards the statement: "*Countries are relevant content across students' subject disciplines*".

Figure 15. Bar chart showing teacher attitudes to the statement: "*Lesser known countries (i.e. Guyana and Tanzania) facilitate discussion*".

Figure. 16. Teacher attitudes to the question: "*How likely are you to use the test again?*"

Figure. 17. Student attitudes towards the statement: *I would like to do the test again*

Table 1. Questions and rationale for EAP teacher questions.

Table 2. Questions for EAP teachers: Overall test evaluation.

Table 3. Questions for EAP teachers: Evaluation of test items.

Table 4. Questions for EAP teachers: Attitudes towards the false accusation.

Table 5. Questions for EAP teachers: Attitudes towards Question 3.

Table 6. Questions for EAP teachers: Conclusion.

Table 7. Questions for EAP students: Introduction section.

Table 8. Questions for EAP students: Overall test evaluation.

Table 9. Questions for EAP students: Individual item evaluation.

Table 10. Questions for EAP students: Conclusion.

Table of Contents

Chapter 1. Introduction	1
Chapter 2. Literature review	
Introduction.....	4
2.1 EAP Context.....	4
2.2 Speaking in EAP.....	5
2.2 i. Group discussions.....	6
2.2 ii. Oral presentations.....	7
2.2 iii. Informal speaking in EAP.....	7
2.3 Testing.....	8
2.4 Testing in EAP.....	9
2.5 Speaking tests in EAP.....	12
2.6 Diagnostic tests in EAP.....	12
2.7 Summary.....	13
Chapter 3. Methodology	
Introduction.....	15
3.1 Background context about subjects.....	15
3.2 Research questions.....	15
3.3 Test design and rationale.....	16

3.4 Theoretical construct definition.....	16
3.5 Operational construct definition.....	17
3.6 Grading	18
3.7 Test procedure.....	20
3.8 Pretesting.....	23
3.9 Research instruments.....	23
3.10 Distribution.....	24
3.11 Questionnaires.....	24
i. Questionnaire 1: EAP teachers.....	25
ii. Questionnaire 2: EAP students.....	29
3.12 Validity & Reliability.....	
.....32	
3.13 Data analysis.....	33
3.13 Ethical considerations.....	
33	

Chapter 4. Results and discussion

4.1 Introduction.....	34
4.2 EAP Teachers.....	34
4.3 EAP Students.....	35
4.4 Overall test evaluation.....	35
i. Validity.....	35

ii. Reliability.....	37
iii. Authenticity.....	37
iv. Impact.....	38
4.5 Evaluation of specific test items.....	39
4.6 Countries as content.....	42
4.7 General attitudes.....	45
Chapter 5. Conclusion.....	49
Reference list.....	53
Appendix 1: Test pack (test paper, country profiles, notes for examiner and mark sheet)	57
Appendix 2: Questionnaire for EAP teachers.....	77
Appendix 3: Questionnaire for EAP students.....	81
Appendix 4: Changes made to questionnaire (EAP teachers).....	84

Appendix 5: Changes made to questionnaire (EAP students).....87

Appendix 6: Email distributed on BALEAP mailing list asking for
participants.....89

1. Introduction

Many higher education institutions' internationalisation strategy has meant students from all over the world now study at English medium universities (Alexander, Argent & Spencer 2008: 5). In the UK, international student numbers have been on an upward trajectory in recent decades: According to the British Council (2013), 198,065 international students were enrolled in UK higher education in 1996 compared to 445,870 in 2013. Alongside the UK, the USA and Australia also rank highly as destinations for international students (UNESCO 2016) and the number of English medium instruction (EMI) courses throughout the world has seen a dramatic rise over the past decade (Dearden 2014).

In addition, the vast majority of academic publishing materials, from textbooks and online resources through to academic journals, are written in English (de Chazal 2014: 29).

English for Academic Purposes (EAP) has emerged as the language teaching profession's response to this. EAP's increase in popularity and professionalism has led to it becoming a "global phenomenon" (de Chazal 2014: 3) and a "multi-million-dollar enterprise around the world" (Hamp-Lyons 2001: 126).

Many EAP programmes look at reading and writing to the relative neglect of speaking (Alexander et al. 2008). This may reflect the modes of language students will need in their subsequent studies. Yet there is acknowledgement that speaking in academic contexts is a skill that is both important and often overlooked (de Chazal 2014; Robinson et al 2001).

Within EAP, testing is also a key issue. There is an ongoing debate in the field as to the most appropriate testing methods. These debates largely focus on the heterogeneous nature of EAP contexts; different institutional requirements and subject disciplines are just two factors which make designing tests that are useful in different contexts a major challenge.

Tests in EAP are carried out for different purposes. Achievement tests show a student's ability and proficiency tests assign students a global grade or score. These are known as summative tests; they make a judgement about students.

Formative testing, meanwhile, is often carried out during a course of study and is concerned with giving feedback to students on their progress. This can help students by giving them

feedback on their progress. They also help teachers by giving them feedback on both the efficacy of the course and their students' progress.

Testing is an area in which EAP teachers should be well versed (BALEAP 2008). Alexander et al. (2008: 325) argue formal formative testing should be “a regular part of EAP classroom routine”. Evidence supports the notion that locally developed tests are strong in terms of both validity and authenticity and there is currently a shortage of such tests (Schmitt & Hamp-Lyons 2015). As a result, it seems plausible to suggest there is room for more ideas on testing within EAP.

This dissertation looks at attitudes of teachers and students towards a self-designed diagnostic test of speaking for EAP. It is designed for students for whom English is a second language who are studying, or about to start studying, at an English medium university.

The test design is based on my own experience as a language teacher, as a postgraduate student in the UK studying alongside a mostly international cohort and on the ample literature on testing from both a practical point of view (e.g. Heaton 1994; Underhill 2004; Davies 1990; Hughes 2003) and a theoretical one (e.g. Bachman 1991; Weir 2005; Lazaraton 2002).

The speaking test was designed with the aim of producing a test that was useful in the sense that it would be a useable EAP assessment tool. Speaking is acknowledged as a “difficult skill to test” (Heaton 1994: 88) and, at present, there is a lack of what constitutes a ‘good’ test of speaking, especially within EAP (de Chazal 2014). Therefore, the test hopes to contribute to this debate.

The test was designed and then initially piloted with students. This revealed some problems of varying severity to be remedied. The refined version was then distributed to EAP teachers throughout the world and administered to students. Each group then gave their feedback through questionnaires. This included teacher and student evaluation of the test in terms of its usefulness as defined by Bachman & Palmer (2009).

As Hughes (2003) posits, it is important the end users of a test, that is, the teachers and students, enjoy using it. With that in mind, the questionnaires aim to address the following research questions:

1. What are EAP teachers' attitudes towards the test?
2. What are EAP students' attitudes towards the test?

This dissertation starts by reviewing relevant literature for testing and EAP (cf. Chapter 2). It then goes on to look at the rationale for the test design and a takes critical overview of the research methodology used to gauge students' and teachers' attitudes (cf. Chapter 3). Finally, results of the questionnaires are analysed in order to reveal the range of attitudes towards the test (cf. Chapter 4). These attitudes will reveal how far along the road the test is to becoming a useful, formative EAP testing tool.

2. Literature review

Introduction

This literature review provides a brief overview of theory and research relevant to testing practices in EAP. Particular attention is given to speaking within EAP and the role of formative classroom based testing methods.

Firstly, the contemporary EAP context is discussed in relation to its main areas of focus and some of its differences to general language teaching.

Secondly, speaking is framed within this wider EAP context. Attention is paid to the construct of EAP speaking, and how this translates to the classroom and testing environment. In particular, focus is given to the requirements at postgraduate level, given that the subsequent test has been designed with postgraduate students in mind.

Finally, testing practices within language learning are then discussed with reference to diagnostic tests, before a more detailed look at testing within EAP.

2.1 EAP Context

EAP originated on the fringes of ESP, but has emerged to become a major force in language teaching today which is both “theoretically grounded and research informed” (Hyland 2006: 4).

Despite being part of the English Language Teaching (ELT) “tree” (Hutchinson & Waters 1996: 17), several authors point out EAP is a branch somewhat apart: It is considered purposeful, high stakes and highly contextual (Hyland 2006; Alexander et al. 2008; Jordan 1997; Hamp-Lyons 2001; de Chazal 2014). It is purposeful in that students have a defined purpose for studying EAP; it is the medium through which they will pursue their subsequent studies. EAP is also high stakes; students often need to pass a test in order to gain entry to their chosen course and, lastly, it is highly contextual. This means that despite some commonalities, the nature of EAP differs across institutions and countries.

EAP’s seemingly straightforward concerns with “the communication skills which are required for study purposes in formal education systems” (English Teaching Information

Centre (ETIC) cited in Jordan 1997: 1) belie what Hyland (2006) claims are the myriad ways that EAP seeks to engage learners in a critical understanding of the varied contexts and practices of academic communication. In practice, this means that in addition to the traditional four language skills of reading, writing, listening and speaking, EAP is concerned with academic literacies. These include critical thinking, study skills and learner autonomy, among others (de Chazal 2014). A further level of complexity lies in the extremely wide range of subjects that may be studied by EAP students (de Chazal 2014: 7). This leads to disciplinary differences which influence how knowledge is expanded and communicated (BALEAP 2008: 3). Subject disciplines may have their own norms of communication. As such, they are considered academic genres; academic writing for finance is very different to academic writing for, say, civil engineering. These differences are exemplified by the extensive research carried out in the field of genre analysis (e.g. Swales 1990). These differences contribute to the difficulty in pinning down a definition of English for General Academic Purposes (EGAP) which can cover these diverse contexts.

While reading, writing, listening and academic literacies are important and deserve attention, academic speaking is widely acknowledged as both essential for overseas students on a personal level (Jordan 1997; Kennedy & Boltho 1996) and an academic one (Alexander et al. 2008; de Chazal 2014). On arrival in a new country, international students often require speaking and listening skills first, but these may have received less attention than reading and writing in their previous language learning (Alexander et al. 2008).

Despite this importance, however, several authors (de Chazal 2014; Robinson et al 2001) have noted the lack of research on speaking within EAP, as greater focus is often given to reading, writing and academic literacies.

2.2 Speaking in EAP

Overseas students have a range of communication needs in academic contexts. Jordan (1997: 45) argues that expressing themselves in speech is one of the biggest difficulties facing many international students. Hence, there is a strong need to develop this part of EAP students' language repertoire.

In particular, Alexander et al. (2008: 235) posit higher education students require "pragmatic competence", that is, the ability to use their speaking ability effectively to achieve a specific purpose. While these specific purposes for communication vary with discipline, participating in group discussions, asking questions, delivering short talks and using informal language for

a variety of purposes emerge as highly relevant to postgraduate study (Jordan 1997; Alexander et al. 2008; de Chazal 2014). With this in mind, a brief overview of the main purposes of academic speaking seems appropriate.

i. Group discussions

Small or large group discussions are intrinsic to the experience of being a postgraduate student. Although, as Jordan (1997: 197) notes, there is disagreement as to what constitutes a seminar. However, most seminars (or tutorials) involve group discussions. These comprise “an integral part of learning and teaching” (de Chazal 2014: 251). In seminars, all students are “expected to make a contribution and respond to contributions from others” (de Chazal 2014: 253), so speaking is a prerequisite. This is reflected in research on needs analyses of EAP students, which has highlighted the importance of “holistic, interactive speaking tasks such as small group oral discussions” (Robinson et al 2001: 348).

International students often find academic speaking their most challenging task (Alexander et al. 2008: 236). A study by Cross & Hitchcock (2007), however, suggests context plays a big role here, at least for Chinese students who make up a considerable part of the international study body. The study revealed third-year undergraduate Chinese students at the University of Portsmouth who were reticent in face to face group discussions, but made contributions that were “lively and insightful” (Cross & Hitchcock 2007: 6) in a computer mediated class environment. This willingness to contribute factor is often lacking in face-to-face encounters. The researchers concluded willingness to contribute, an important part of speaking, is often a case of context rather than an innate or cultural trait.

The conventions of group discussion are often not made explicit and may be very different to overseas students’ previous learning experiences. So EAP tutors are in a position to prepare students for such discussions by helping them gain awareness of the purposes of the discussion and develop participation strategies (Alexander et al. 2008); a key part of speaking.

At postgraduate level, seminars often contain divergent tasks, that is, topics about which agreement is not necessarily the aim (Alexander et al. 2008: 236). Postgraduate students are expected to become members of their discipline’s discourse community, and group discussion is a key part of this.

A key component of the group discussion is asking questions. They allow clarification, an interpretation check, a digression or even a challenge (Jordan 1997: 193). Questions also

contribute to general discussions by providing feedback to the speaker and showing the listeners' level of understanding. Questions are also valued contributions in themselves regardless of their function. As Jordan (1997) notes, students should be sensitised to the idea that in many educational contexts a positive value is attached to the notion of the person who asks questions.

ii. Oral presentations

Short talks and presentations are deemed essential for EAP students, especially those at postgraduate level who will be asked at some stage to give a short talk on some aspect of their studies (Jordan 1997: 201). In addition, they offer good practice at improving fluency. Since giving talks also exposes students' "core fluency" for teachers to observe and rectify, they offer diagnostic items for language testing (Jordan 1997, p203). That is, teachers can hear aspects of students' vocabulary, pronunciation and grammar in an uninterrupted flow.

Dudley-Evans & St Johns (1998) posit presentations constitute the most important aspect of EAP speaking because they often form part of a programme's assessment. Both de Chazal (2014) and Alexander et al. (2008) argue that EAP students should be well versed in the two main parts of presentations: content and delivery. These two must be balanced in order to achieve the presentation's aim, and clear criteria exist for achieving this.

iii. Informal speaking in EAP

The ability to use informal language when speaking is also relevant to EAP students (Jordan 1997: 193).

There can be a misapprehension that, by definition, EAP cannot be concerned with social skills, yet it is clear that EAP students need to function socially as well as academically and this must be taken into account (Kennedy & Bolitho 1996: 117). This is undoubtedly necessary for international students arriving in an English medium country. Registering with the doctor, opening a bank account and integrating socially are all important, for instance (Alexander et al. 2008: 234). This informal language can also be used in academic contexts. Group discussions may use a more informal register and lecturers frequently digress. Furthermore, in supervisory meetings students may frequently be required to take the lead in thesis and other supervisory sessions with tutors (Alexander et al.2008: 236).

2.3 Testing

The terms assessment and testing are often used interchangeably in the language teaching literature (de Chazal 2014: 292). However, de Chazal (2014: 292) points out that assessment is a broad term which refers to all types of student evaluation, of which testing is one form.

Language tests can be a valuable tool. Information about people's language ability is often very useful and sometimes necessary for providing evidence of teaching and learning, and providing feedback on the effectiveness of a teaching program (Bachman & Palmer 2009: 8). Yet while they can serve such pedagogical functions, their main function is "to measure" (Bachman & Palmer 2009: 19). Hence, a common yardstick of ability is often required to make meaningful comparisons between students (Hughes 2003: 4).

When looking at test development and use, the first question that often arises is whether the test is any good. Bachman and Palmer's (2009) seminal work aims to answer this by evaluating whether or not a language test is 'useful'. In their eyes, six criteria must be met for a test to be considered useful. Each quality, however, may be more or less important depending on the purpose of the test.

- *Reliability* is a test's consistency in yielding results. The same candidate should be graded similarly by different examiners, for example. While some aspects of reliability are outside control, such as candidates' moods on test day or their background knowledge, minimising variation in task type and delivery may be controlled in order to enhance reliability. In other words, students should not be disadvantaged in any way due to setting or examiner (Alexander et al. 2008: 315). A high-stakes proficiency test aims for strong reliability as it will be administered to many candidates in different locations. Meanwhile, a low stakes classroom based test may not need such reliability considerations if it is administered to a relatively small group of candidates.
- *Construct validity* refers to the notion that a test should measure the types of abilities that are intended for measurement, and nothing else (de Chazal 2014: 297). This is particularly relevant in an EAP context where students are learning language for a specific purpose (academic study) and therefore should not be assessed on areas not related to this.
- Test *authenticity* relates to how realistic or authentic test tasks actually are. This factor has been found to affect test takers' performance as tests with greater authenticity can promote a positive affective feedback response (Bachman & Palmer 2009: 24).
- The *interactiveness* of a test refers to how far candidates are engaged by the test. A test that requires a test taker to relate his or her knowledge to the topical content of the test is likely to

be more interactive than not. This could also be thought of as the degree of difficulty of the test for the students.

- *Impact* means the effect that test results have on the individual, the institution and society. A test can have pedagogic consequences often referred to, sometimes negatively, as ‘wash back’ whereby teaching is focused on exam strategies rather than language skills per se. Test impact also refers to decisions that arise from its result: entry to university, for instance.
- Test *practicality* relates to the resources needed to administer it. This encompasses human resources, materials needed, time and any financial cost.

As Alexander et al. (2008: 314) claim, these criteria are a good way to rate the usefulness of a test for a particular purpose.

2.4 Testing in EAP

Davidson & Cho (2001) chart the course of EAP testing procedures from the late 1940s to the present and note how testing approaches have generally reflected the views of language learning at the time.

Large-scale proficiency tests, such as IELTS and TOEFL, are useful for determining entry criteria to university courses. This is one reason why de Chazal (2014: 291) calls EAP tests “essential”. While this function is acknowledged, several researchers claim these tests often fail to instil the pragmatic competence needed for successful academic study. A study by Ockey et al (2015) compared university students’ TOEFL scores with performances on various academic speaking tasks. They found TOEFL was a good indicator of academic oral ability in terms of grammar, pronunciation and lexis, but the test was not such a good indicator of interactional competence, descriptive skills, and presentation delivery ability. De Chazal (2014: 15) points out that among EAP tutors there is a “strong belief” that success in general proficiency tests does not enable students to function well at university, as they lack the academic literacies (e.g. note taking, referencing, brainstorming in group discussions) to do so.

One possible reason for this is that EAP has a broader construct than normal oral assessment (Schmitt & Hamp-Lyons 2015). This construct means students should apply their language ability to completion of academic tasks. In particular, authentic tasks should be used, which often involve integrated testing: Combinations of speaking, listening and reading, for example. While most general language tests may separate skills into ‘speaking’ and

'listening' sections, de Chazal (2008: 241) argues listening and speaking are "inseparable" in an EAP context.

Douglas (2000: 19) extols the virtues of being as specific as possible in testing to reflect the importance of context and precision when using language for specific purposes. Similarly, there is broad agreement in the literature that EAP tests should assess the abilities that are actually required for a particular level of tertiary study (Brindley & Ross 2001; de Chazal 2014; Alexander et al. 2008; Douglas 2000; Bachman & Palmer 2009). In other words, any test should have a high level of construct validity. Hughes (2003: 1) is blunt on the subject: "Too often language tests have a harmful effect on teaching and learning, and fail to measure accurately whatever it is they are intended to measure". This also means research into EAP instruction and testing in one programme cannot be generalised to others due to the influence of local factors, such as students' L1, cultural background and institutional constraints like programme aims and resources. However, general questions are often transferrable to comparable EAP contexts (Robinson et al 2001: 348).

Schmitt & Hamp-Lyons (2015: 3) argue that the construct in EAP assessment is both "under-defined and under-theorised" with research focusing on large-scale exams (e.g. IELTS and TOEFL) to the neglect of locally developed in-house EAP assessment. Indeed, they argue that EAP practitioners are better placed to define the construct than language testers who design large-scale standardised exams. A study by Jameson, Wang & Church (2013) found a test developed in-house for EAP placement proved more useful than a commercially available product, as the former was more practical in terms of cost and time, and also more authentic as it was based on the local context.

As a result, there is growing awareness that classroom based tests should feature in EAP contexts. O'Sullivan (2011 cited in Schmidt & Hamp-Lyons 2015: 5) posits that "in almost all circumstances local tests are more likely to allow us to make valid assumptions about test takers". Alexander et al. (2008: 321) argue the EAP classroom is a "rich social space" that allows a more authentic interaction than in high stakes exams. Therefore, classroom tests can focus on validity, authenticity and impact. Meanwhile, global exams tend to focus on reliability and practicality in order to cater to large numbers of candidates (Alexander, Argent & Spencer 2008: 321).

Yet localization needs to be accompanied by assessment literacy on the part of instructors: Schmitt & Hamp-Lyons (2015: 5) believe that testing literacy of teachers is "of great concern" and argue for a closer interaction between EAP teachers and language testers as the

skills and knowledge of both parties seem mutually beneficial. This is echoed in the BALEAP competency framework (2008) which states EAP teachers should not only be *au fait* with the links between assessment, teaching and learning, but also be able to “select appropriate modes of assessment and also design assessment tools for language and skills related to EAP speaking, including integrated skills” (BALEAP 2008: 9). Formal testing should be a regular aspect of EAP classroom routine as it can give students a sense of purpose in their learning and provide feedback for teachers for fine-tuning the course.

A major theme which emerges in EAP assessment is content validity, that is, the amount of subject specificity in texts and materials. Some have labelled EGAP a misnomer (Hyland 2006: 12), instead favouring an ESAP approach. There is broad agreement that seminar conventions vary depending on the discipline and the aims of the seminar leader (Kennedy & Boltho 1996; Jordan 1997). Meanwhile, others maintain that many language skills overlap different subject areas, but tutors should still aim to find out more about their students’ disciplines (de Chazal 2014: 245). An EAP tutor should therefore look at specific seminars and teach the appropriate strategies, where possible. Hyland (2006: 4) notes that research has undermined the notion of “a single monolithic academic English” and disciplinary variations are now rightly acknowledged. This also raises questions about content validity in tests; the amount of subject specificity that should be used in texts and tasks. This point will be discussed in due course in relation to the test.

As a result of this, it appears there is a demand for locally developed formative EAP tests and that currently there is a lack of tests available. As the literature suggests, there are differing contexts in EAP, yet with some commonalities and a strong agreement that locally developed tests have high degrees of validity and authenticity. It is possible, then, that a classroom test for one context could contain some elements that are useful to other contexts or contribute to the development of other tests.

2.5 Speaking tests in EAP

While speaking is seen as an important part of any language test, it is acknowledged as “an extremely difficult skill to test” and, what’s more, test results often prove “too complex to allow reliable, objective analysis” (Heaton 1994: 88). Like other areas of assessment, tests of speaking are in constant evolution (Davidson & Cho 2001) and at least some of the difficulty and subjectivity to which Heaton (1994) alluded can be reduced by creating a specification of content along with standard procedures for marking and grading the test (Council of Europe 2001). However, as Alexander et al. (2008) point out, all language tests make compromises.

Communicative ability is an abstract concept that cannot be measured directly. Instead, tasks may elicit language behaviour which can be measured. Yet even these measure language behaviour in a test situation, not an authentic one.

Several writers (Hughes 2003; Heaton 1994; Underhill 2004) argue that the interview is by far the best assessment of speaking. Although as Hughes (1996: 104) notes, an interview between tester and candidate is asymmetrical, and the candidate is unlikely to take the initiative during the interaction. A move to remedy this is to assess peer interaction by having two or more candidates in a test (Hughes 2003: 104).

There is also a belief that assessment methods should mimic instruction methods as much as possible. A study by Robinson et al (2001) found university students of a similar calibre performed differently in an academic speaking test depending on the method of instruction used.

2.6 Diagnostic tests in EAP

An important distinction is made between formative and summative testing. The results of the former are used by teachers to improve instruction while the latter often constitutes the end of semester or course assessment (Brindley 2001: 137). As Alexander et al. (2008: 304) put it “formative is developmental, summative is judgemental”. Formative assessment is often low stakes and its results’ impact is limited, whereas summative assessment often carries with it high stakes; results could dictate whether the test taker is admitted to university or given a job promotion, for example.

Diagnostic tests are formative and aim to identify the taker’s strengths and weaknesses, and indicate what learning still needs to take place (Hughes 2003: 15). Students benefit from this but so do teachers: diagnostic tests give feedback about the effectiveness of teaching and the programme itself. Hughes (2003) points to the lack of good diagnostic tests in English language testing, which he describes as “unfortunate”.

A study by Fox (2009) looked at a top-down policy which made proficiency tests the requirement for placement at a Canadian university. As a result, the institution developed diagnostic tests which allowed teachers to focus on students’ weaknesses and build learner profiles. The use of diagnostic tests here proved highly successful as they moderated the effect of the policy and encouraged curricular renewal.

2.7 Summary

In summary, the literature seems to reveal that global tests are useful in EAP, but also have their limitations. To address this, local or classroom based tests should be developed to cater for one part of the formative aspect of assessment. EAP teachers should be 'assessment literate' to deal with this. Of all the necessary skills within EAP, there seems to be a lack of research into academic speaking, and particularly into speaking assessment. Therefore, it could be suggested there is a niche within research on EAP speaking tests.

3. Methodology

Introduction

This study involves researching the development of an EAP speaking test and gauging attitudes of teachers and students towards it. After the initial test design, the test was piloted (cf. Section 5) in order to evaluate its design. From this, a second version of the test was produced and piloted. This second pilot involved observing how another examiner, an experienced language teacher, would administer the test which would highlight any shortcomings in the instructions, along with any issues with the test itself. The test was administered to 12 current postgraduate students from a range of non-English speaking countries. They were studying subjects such as business, sports management and TESOL at a UK university. A third version of the test was then developed and distributed to EAP teachers. The test was distributed to EAP teachers directly and advertised through the BALEAP email list. It was hoped that EAP teachers would look at the test, perhaps use it with their students, and then give their feedback on it via a questionnaire. The test was also administered to pre-sessional and in-sessional postgraduate students who were then asked to complete a questionnaire on their attitudes towards the test. This process is summarised in the timeline below:

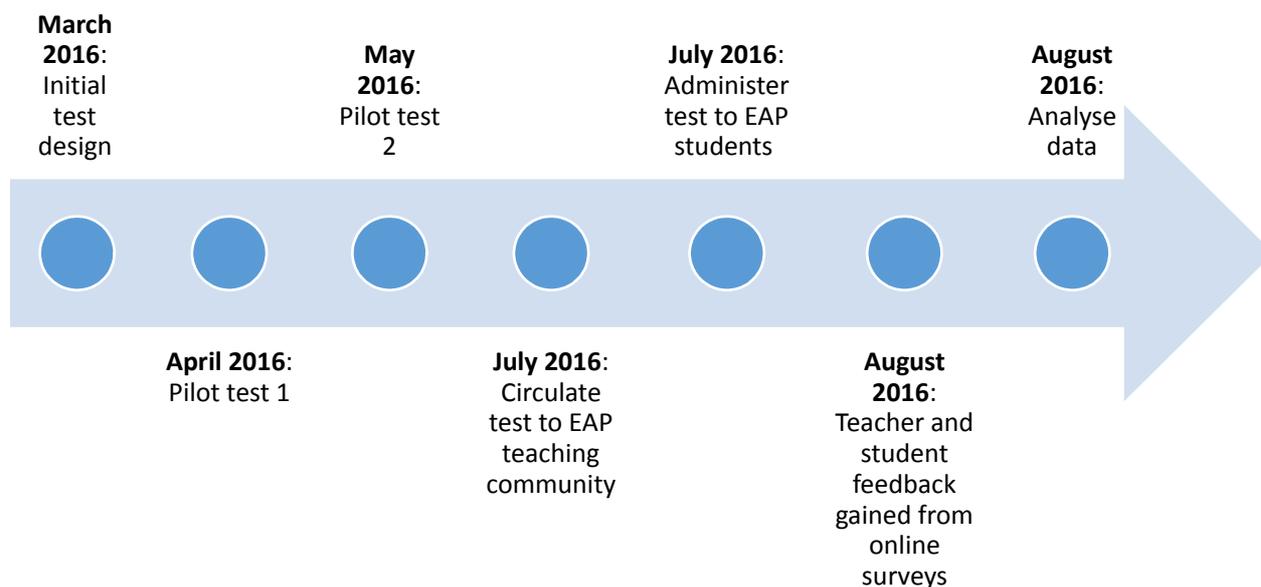


Figure 1. Research timeline showing the order of test development and data gathering

3.1 Background context about subjects

The research concerned gauging the attitudes of both teachers and students towards a test so the primary subjects were the end users of the test: the teachers and students. In this case, the target student population were postgraduate students for whom English was a second language. They were studying, or about to study, a postgraduate degree at an English medium higher education institution in the UK. They were from China, Japan, Cyprus and Greece. There were 5 males and 8 females.

EAP teachers came from countries such as the UK, USA, Hungary and Germany. They worked in the UK, USA and Germany.

3.2 Research questions

The research questions focused on end user attitudes towards a self-designed EAP diagnostic speaking test and are stated below:

1. What are EAP teacher attitudes towards the test?

2. What are EAP student attitudes towards the test?

It is hoped that student and teacher reactions to this test can be gained in this way. This is an essential step in the process of test development; as Underhill (2004: 105) notes, both parties' enjoyment of using the test is intrinsic to its efficacy.

3.3 Test design and rationale

The test which is the object of this research (hereafter referred to as 'the test') (cf. Appendix 1) was self-designed based on my previous teaching experience and my experience as a postgraduate student working alongside an international cohort at a UK university. The initial step was to design the construct, that is, what is to be tested, by close scrutiny of the language demands of postgraduate students in higher education.

Bachman and Palmer (2009) argue that test design has three important steps. Firstly, the construct needs to be defined theoretically. In other words, the material that is to be tested. Secondly, the construct needs to be defined operationally. That is, how the construct will be tested; this concerns tasks, test items and how the test will be administered. Then a method must be established for quantifying responses to tasks in order to give candidates a mark for the test. Lastly, an outline of the test procedure and rationale is given.

3.4 Theoretical construct definition

This first step in test construction is what Hughes (2003: 48) calls the "statement of the problem" and Weir (2005) refers to as the *trait*. In other words, what exactly one wants to know about the candidates and the purposes for which this information will be used. Once this problem has been stated concisely, clear steps can be taken to answer it.

The test is a diagnostic test of speaking for EAP students. So Hughes's (2003) 'problem' in this sense is identifying areas of weakness in English as a Second Language (ESL) speakers' academic speaking skillset. This identification would give feedback to students and teachers. Students would then be aware of areas in their speaking repertoire which require improvement. Meanwhile, teachers receive feedback on the efficacy of a speaking course, and provide them with guidance on what to focus on. So as well as evaluating the effectiveness of an individual student's learning, assessment they can also evaluate the broader effectiveness of the programme and its teaching (de Chazal 2014: 291).

Items in the test measure types of language and tasks that candidates will be expected to use and perform at university, such as participating in group discussions, asking questions, making presentations and contributing to discussions based on reading and listening input. As Weir (2005: 75) argues “test tasks should be representative of the target situation” and, as a result, all of these elements are present in the test.

However, the EAP construct is very much determined by the local context; the students, their subject discipline, level of study, language requirements and institutional factors. As such, a ‘one-size-fits-all’ approach is not suited to EAP and is one likely reason global proficiency exams have been criticised for not suitably preparing EAP students for academic life (e.g. Alexander et al. 2008; de Chazal 2014; Schmitt & Hamp-Lyons 2015).

That said, the test aims to appeal to EAP teachers in different contexts by adopting a flexible framework that is adaptable to local contexts. As Robinson et al (2001: 347) argue, there are several general factors that are transferrable across comparable EAP environments. With that in mind, a ‘notes for the examiner’ document has been drafted (cf. Appendix 1) which gives instructions to interlocutors in administering the test, but also emphasises a flexible approach in that content, time or manner of delivery should be modified as appropriate.

Now the problem has been stated, a set of specifications must be made for the test. This includes decisions on content, format, timing, criterial levels of performance and grading (Hughes 2003: 49), which will be discussed in the following section.

3.5 Operational construct definition

The operational construct covers the set of skills the test aims to measure, such as language functions, the types of text or materials that should be used, who the audience of the test will be and the topics focused on.

Topic is particularly important in EAP. Douglas (2000: 19) argues that any ESP test, and EAP falls firmly within ESP, should derive its content from the target language use situation in order to be authentic. A key feature of the test was the use of countries as content. As a test aimed at EGAP students, it is suggested that country information is broadly relevant across subject disciplines. The use of countries as a topic is one part of the test about which EAP teacher opinion will be sought (cf. Question 7 in Appendix 2).

The format, or what Weir (2005) calls the *method*, meanwhile, concerns test structure; how it is delivered along with item types and elicitation procedures. Test tasks were designed so as to mimic as much as possible similar tasks that would be required of university students.

Along with this, setting is also important. Weir (2005: 56) argues the physical location of a test should be as similar as possible to the setting where the language will eventually be used. Hence, the test should be administered in a classroom or similar environment. The test's rubric should also be clear to candidates as well as "brief, candidate-friendly and simple" (Weir 2005: 57).

3.6 Grading

This is the final stage in Bachman & Palmer's (2009) three steps; quantifying responses to tasks in order to give candidates a mark for the test. Weir (2005) argues that a test's scores form a key part of its validity claims. Test scores should be fair representations of what candidates can do which requires strong construct validity and an adequate grading system.

Grading speaking tests is mired in difficulty, largely due to the subjective nature of their marking and the consequent effects on the test's reliability. Several researchers reject traditional grading systems. Moss (2003) argues against giving grades in formative assessment, suggesting it takes too much time to develop appropriate descriptors for each formative task and to assign grades. Meanwhile, Lazaraton (2002) shirks the idea of quantitatively grading tests, focusing instead on qualitative means such as conversation analysis. She sees assessment as a fundamentally social act which requires a method of grading that reflects this.

However, an objective scoring criteria based on realistic expectations of student performance and needs can give the test some degree of objectivity (Heaton 1994: 100) and can also be motivating for students (Alexander et al. 2008: 305). Furthermore, it is generally agreed (Underhill 2004; Heaton 1988; Hughes 2003) that grading should penalise language errors that impede meaning more heavily than those that do not detract from the overall message.

Since the test is diagnostic, priority is given to the type of feedback that results from it rather than a numerical grade. The grading system was developed with this in mind. Examiners use two mark sheets, one for each candidate, and carry out impression marking, that is, "subjective judgement on a range of criteria" (Underhill 2004: 101). Subjective marking of this sort is advantageous in acknowledging the complexity of language (Council of Europe 2001:188).

A criterion-referenced grading system is used in the test. This measures students' performance against a benchmark standard (Alexander et al. 2008: 311). Candidates are awarded a score from 0 to 4 for each of the criteria. 0 indicates no evidence of the criteria and would be awarded if, for example, the candidate remained completely silent on being asked a question; there would be nothing to assess. A 4 would be awarded for a masterful display of the criteria, akin to a masterful display for the given level. The goal should be realistic expectations of what learners can achieve at a particular stage of their learning (Heaton 1988: 100). Bachman & Palmer (2009: 212) argue that any scale should have a zero and mastery dichotomy, "irrespective of whether any test takers are at these levels". The number of notches between these extremes was dictated by need; too many may be impractical for an examiner during a test.

Grading criteria include those focusing on language (vocabulary, accuracy and fluency) and then categories based on performance criteria for each question. At the end, interlocutors record anything that stuck out as a potential weak point. This is then given to the candidate and forms the basis for further learning and teaching, providing a focus on areas which require improvement.

As mentioned, the test has been developed with the expectation that certain parts may be modified to fit the local context. Similarly, the grading criteria should be subject to moderation sessions in which examiner/teachers discuss their decisions about grades. Alexander et al (2008: 335) claim that this is essential due to the myriad "beliefs, attitudes and experiences...which inevitably mean teachers give weighting to different parts of any grading criteria".

Despite their use, rating scales also have drawbacks. In particular, raters may interpret a scale differently, have different standards of severity and subjectivity may arise over issues not on the scale; previous knowledge of candidates' speaking performance or empathy with views on issues, for example (Bachman & Palmer 2009: 221). This influences the reliability of the test. In an attempt to address these issues, a detailed scoring criteria was drafted (cf. 'Notes for examiners', Appendix 1).

Scoring is seen as essential to test development (Weir 2005; Bachman & Palmer 2009) and, much like other decisions concerning test development, should be checked by pre-testing and then evaluating results. This was carried out in the test during the first two pilot tests.

A brief summary of the test procedure and its rationale now follows.

3.7 Test procedure

The interlocutor warmly welcomes candidates on their arrival at the test location. Several researchers (e.g. Underhill 2004; Hughes 2003; Heaton 1994) posit that making candidates feel at ease from the start is not only courteous, but contributes to their subsequent speaking performance; candidates often feel anxious at the prospect of an oral test.

Candidates are then given a country profile to read (cf. Appendix 1). This is a short text about a country, either Guyana or Tanzania. Countries were deliberately chosen as a topic for the test as they fulfil criteria set out by Alexander et al (2008: 243) that chosen topics should “require preparation and be able to be discussed with academic rigour”. This notion is supported by Weir (2005) who believes texts should be “suitable in terms of genre, rhetorical tasks, patterns of exposition and at an appropriate level of specificity” (Weir 2005: 75). In particular, content of any text used should be “sufficiently familiar to candidates so they have sufficient schemata to enable them to deploy appropriate skills and strategies to understand the text” (Weir 2005: 75). A text should not be too arcane to be mapped onto a student’s existing knowledge, nor should it be too familiar (Weir 2005: 76). Hence the test uses the lesser known countries of Guyana and Tanzania. Candidates know the framework of a country but may be unfamiliar with the specific countries used in the test. Countries are believed to feature in most academic disciplines in some way and may represent a relevant general topic for EGAP students. However, the argument that countries represent appropriate content for EGAP students is a hypothesis that will be measured in the questionnaire used in the present study (cf. Question 7, Appendix 2).

Candidates are given ten minutes to read the text (approximately 200 words) and told they will give a brief talk about the country as part of the test and then have a discussion about it. They are encouraged to highlight any language or points that they do not understand as they will be invited to ask any questions about the text. This part of the test aims to mimic, albeit in a highly abridged way, what students may reasonably be expected to do in a seminar or tutorial: preparing beforehand by reading about a topic and then engaging in a group discussion which aims to “develop a deeper level of understanding through analysis and negotiation of ideas” (de Chazal 2014: 252). What’s more, candidates may “demonstrate their understanding by being able to reformulate and summarise ideas” (Alexander et al 2008: 237). This task also allows students to deliver a monologue which several writers (de Chazal 2014, Alexander et al 2008; Jordan 1997) state is an important aspect of speaking in EAP. However, Hughes (2003: 109) cautions that a prepared monologue is an unsuitable task for

testing since it is unrealistic and may cause considerable student anxiety. Hence this task aims to strike a happy medium; students are given a short period to think about what they might say, but not long enough to prepare a script.

After they have read the texts, candidates are invited back in to the test room. They are given the relevant student copy of the test and directed towards Question 1. The first item requires both candidates to interact with each other as if they were sitting next to each other and waiting for a lecture to begin. This item is both a role-play which “allows the elicitation of some target functions” (Hughes 2003: 107) such as questions and a peer-to-peer interaction, which avoids the asymmetrical relationship of an interlocutor-candidate interaction. This question allows students to demonstrate their ability from the start, as Hughes (2003: 106) posits “it is especially important to make the initial stages of a test well within the capacities of all reasonable candidates”, allowing them a straightforward start to the test which would reduce any test anxiety they may have. It also aims to encourage candidates to take the initiative in asking questions, which Jordan (1997: 195) argues is a key area of academic speech and one in which students are frequently lacking. Candidates may be using a more informal register to communicate and talking about general topics, rather than strictly academic ones. However, Jordan (1997) and Alexander et al. (2008) point out that adopting a more colloquial register is still an important skill for EAP students that should not be downplayed. Again, the relevance of informal language in an EAP test is a hypothesis that is to be tested in the present study (cf. Question 6, Appendix 2).

Question 2 returns to the *country profile* text students were given to read at the start of the test. Firstly, candidates are invited to clarify any language or other points with the interlocutor. Uncommon words were deliberately used in the texts (e.g. *creole*, *aegis*, *cassava*) in an attempt to galvanise candidates to clarify unknown words; as Jordan (1997: 194) states, clarifications are a key question category for EAP students. After answering any queries, the interlocutor invites the first candidate, Candidate A, to deliver a short *ad hoc* talk about the country, stressing that only information included in the text should be in the talk, and that it should be appropriately structured with an introduction and conclusion. This allows delivery of a short talk, a so called “prototypical academic communicative event” (de Chazal 2014: 251). It also allows examiners to listen to candidates’ speaking ability from a language perspective since monologues of this type expose a student’s fluency in a continuous flow, rather than in a short answer. Teachers may observe and rectify this and so short talks of this nature offer “diagnostic items for language testing” (Jordan 1997: 203).

While Candidate A is delivering the talk, Candidate B is taking notes, having been informed that she/he will be required to instigate a discussion about the country after the talk. De Chazal (2014: 254) claims academic speaking and listening are “two sides of the academic communication coin, in other words, totally inseparable in EAP” so this item aims to incorporate this. After the two-minute talk, Candidate B is invited to begin a discussion lasting two minutes. As stated in the question, this could be a clarification question, stating an opinion or asking a general question; all valid EAP functions in themselves, but the overall goal is to engage in an exploratory discussion of the country with the other candidate.

After two minutes, the interlocutor thanks both candidates and announces that he/she will now join the discussion for a further 3 minutes. Candidates must now interact with each other and the examiner. Alexander et al (2008: 237) point out that EAP students are often not used to a teacher adopting a more informal tone in a group discussion, so this item aims to familiarise students with this. However, the task is similar to the prior one in that the candidates and examiner are engaged in a group discussion to explore meaning. A key aspect of this part of the test is that the examiner delivers a false assertion to the candidates. Candidates would then be marked on their ability to challenge the examiner, deemed an important skill in EAP (Jordan 1997: 193) and showing an element of critical thinking. If corrected, the interlocutor would apologise and thank the candidate for the correction. Again, teacher and student attitudes towards the inclusion of this false assertion in the test will be measured in this study (cf. Question 9, Appendix 2).

The third and final question involves students interacting with each other again. Candidates are given two minutes to discuss who will do what, where they might find the information and when they can meet up and discuss their progress. This peer-to-peer interaction is seen as preferable to that of the interlocutor-candidate one, but as Hughes (2003: 104) warns, one candidate can often dominate the interaction, reducing the other’s opportunities to show his/her speaking ability. This is an inherent problem and one that can only be remedied by hand picking candidates of similar ability and complementary character. As this is a classroom test administered by teacher, this could be feasible. Candidates are given a group assignment which includes a deadline and about 5 or 6 tasks which must be fulfilled. This is one of the “key academic communicative events” to which de Chazal (2014: 245) refers. These consist of two or more students working together on a project and involve “collaboration, negotiation, problem solving and is outcome driven” (de Chazal 2014: 245).

The test includes a group assignment drawn from the subject discipline of English literature. However, a quick glance at the question confirms that no content knowledge is required to complete the task. Similar group assignments from other subject disciplines could easily be drafted.

3.8 Pretesting

Pretesting is an essential step in test construction. As Hughes writes (2003: 51) “the writing of successful items is extremely difficult” and hence pretesting allows some problems to be identified and amended as necessary. Two pilot tests were carried out, which revealed issues of varying severity.

3.9 Research instruments

In order to gauge teacher and student attitudes on the test, two questionnaires were designed. Surveys of this nature allow researchers to obtain “a snapshot of conditions, attitudes and/or events at a single point in time” (Nunan 1992: 140). In addition, questionnaires provide structured data that is relatively simple to analyse (Cohen et al 2011: 377).

The questionnaires (cf. Appendices 2 & 3) used a combination of open and closed questions. Open questions are useful for site-specific, qualitative, less structured, word-based answers and useful for specific situations. They also provide rich detail in giving respondents freedom to say anything. Yet Dörnyei (2003: 47) cautions that open questions offer only superficial, short answer possibilities since questionnaires “are not the right place for essay questions”. However, they still offer illustrative quotes and could flag up issues not anticipated, particularly relevant in the present investigation of test administrators’ and takers’ attitudes.

Closed questions offer only a limited range of responses and are useful where measurement is sought and hence are more quantitative (Cohen et al. 2011). Their major advantage is that their coding and tabulation is straightforward and leave little room for rater subjectivity (Dörnyei 2003: 35). This mixture of open and closed questions constitutes what Cohen et al. (2011: 382) call “a powerful tool”. This offers some quantifiable data along with richer, qualitative insights into attitudes.

While their utility is widely acknowledged, construction of valid questionnaires is “much more complex than might at first be thought” (Nunan 1992: 142). Decisions were taken on question content, question wording, formatting, anticipated responses and the sequence of

questions (Cohen et al 2011: 379). An established way of establishing whether a questionnaire is likely to fulfil its purpose is to carry out a pilot.

A pilot was carried out with three respondents; one teacher and two students. This revealed some ambiguities in the wording of questions and in the sequencing of items which resulted in some alterations being made (cf. Appendices 4 & 5).

The questionnaires started with an introduction. This is an essential preamble which contained such information as the name and contact details of the researcher, the purposes of the questions and assured respondents' anonymity.

3.10 Distribution

The test was made available to EAP teachers with the intent they would look over the test and/or use it with their students. Subsequent feedback would then be collected via an online questionnaire. The test was sent directly to several universities in the UK and an email was sent out on the BALEAP mailing list (cf. Appendix 6), which has 1,154 recipients involved in EAP provision in the UK and abroad.

In addition, I administered the test to 14 pre-sessional and in-sessional postgraduate students. After the test they completed a questionnaire which would gather their attitudes towards the test.

3.11 Questionnaires

Questionnaire items will now be discussed along with their rationale in addressing the research questions. Given the following research questions:

1. What are EAP teacher attitudes towards the test?
2. What are EAP student attitudes towards the test?

3.12i Questionnaire 1: EAP teachers

The questionnaire was divided into three main parts. With the exception of Question 4, all questions were optional.

The following tables illustrate the questions asked in the questionnaire for teachers along with a rationale for each question.

Table 1. Questions and rationale for introduction part of EAP teacher questionnaire.

Question	Rationale
1. Please indicate your nationality.	To obtain demographic information required for background context of respondents.
2. Please indicate the country in which you currently work.	
3. Please indicate your gender.	
4. Did you use the test with your students?	To gauge number of teachers who have looked at test compared to those who have used it.
a. If yes, please indicate if the students were: pre-sessional undergraduate / pre-sessional postgraduate (non PhD) / pre-sessional PhD / other (please specify)	To indicate level of students; test was designed with postgraduates in mind.
b. If yes, how many students did you administer the test to?	To gather data on the amount of times the test was used.

The second part focused on test usefulness according to Bachman & Palmer's (2009) six criteria of test usefulness previously defined (reliability, validity, impact, practicality, authenticity and interactiveness). These criteria have been supported by other authors (e.g. Alexander et al 2008, McNamara 2001, de Chazal 2014) as being good criteria to rate a particular test's worth and hence they were used in determining attitudes towards the test. The following statements were given along with a scale of agreement: *strongly agree/agree/disagree/strongly disagree*.

Table 2. Questions for EAP teachers under the heading 'overall evaluation of test'.

Question	Rationale
5.1 The test measures the types of speaking tasks that EAP students will need in their academic studies.	Aims to measure whether teachers think the test has validity.
5.2 The content of the test is relevant to EAP students.	Aims to gauge if teachers think the test has authenticity.
5.3 The test is easy to administer in the classroom.	Aims to measure whether the test is practical.
5.4 If the test was mandatory, it would require specific test preparation classes.	Seeks to measure the impact of the test. If specific classes were needed, it would suggest there are washback effects.
5.5 Both candidates have the same opportunity to demonstrate their speaking ability.	To measure teacher attitudes towards reliability of the test.
5.6 The test measures students' academic speaking ability.	Aims to measure teacher attitudes to the interactiveness of the test.

The next section dealt with the individual questions in the test, but with an emphasis on Question 2 since this question contained the most novel features, such as the use of countries as content in EAP testing and the inclusion of a ‘false assertion’ by the interlocutor.

The following questions used a 4-point Likert scale of agreement, asking participants to indicate the extent to which they agree with a number of statements.

Table 3. Questions for EAP teachers under the title ‘Evaluation of test items’.

Question	Rationale
6. Question 1 focuses on more informal speech between students. This is appropriate for an EAP test of speaking. <i>strongly agree/agree/disagree/strongly disagree</i>	To gauge attitudes on whether informal language should be tested in EAP.
7. Question 2 uses country profiles as content. Please indicate your general impression of countries as content in the test.	To allow respondents to give their general opinion of countries as content. It was placed before the subsequent questions to avoid ‘priming’ respondents’ answers.
8.1 Countries are relevant content across students’ subject disciplines. <i>strongly agree/agree/disagree/strongly disagree</i>	To measure attitudes to countries in a way that is easy to analyse.
8.2 Lesser known countries (i.e. Guyana and Tanzania) facilitate discussion. <i>strongly agree/agree/disagree/strongly disagree</i>	To test the hypothesis that unknown countries would facilitate an exploratory discussion.
8.3 Uncommon words were deliberately included in the country profiles; this facilitates discussion. <i>strongly agree/agree/disagree/strongly disagree</i>	To test the hypothesis that unknown words would encourage clarification questions.

The next part of the questionnaire contained questions related to the false assertion in the test. Since this was a novel feature, teachers’ attitudes on its viability were sought.

Table 4. Questions concerning the false assertion and containing a Likert scale response option.

Question	Rationale
9. Please indicate your general impression of the false assertion in the test.	An open ended question that aims to collect all opinions on this feature.
10.1 At least one candidate noticed the false assertion. <i>strongly agree/agree/disagree/strongly disagree</i>	To discover how many candidates noticed this feature; if all noticed then it could be too easy; if none noticed perhaps too difficult.
10.2 At least one candidate challenged me	To obtain data on the number of challenges;

after giving the false assertion. <i>strongly agree/agree/disagree/strongly disagree</i>	suggesting whether this feature galvanises a response from students.
10.3 This false assertion is appropriate for an EAP speaking test. <i>strongly agree/agree/disagree/strongly disagree</i>	To gauge whether practitioners in the field think this is valid for an EAP test.

In order to have asked about all three main questions in the test, teachers' attitudes towards Question 3 was also sought.

Table 5. Question seeking attitudes on Question 3.

Question	Rationale
11. Question 3 asks candidates to discuss a group assignment. Please indicate your level of agreement with the following statement: This type of task is relevant to EAP students. <i>strongly agree/agree/disagree/strongly disagree</i>	To gauge attitudes towards the validity of the task type in question 3.

The final part of the questionnaire asks whether teachers are likely to use the test and also contains a general question; as Dörnyei (2003: 50) argues, a short answer question involving a real exploratory enquiry about an issue can be a “motivating and good way to finish off”.

Table 6. Questions in the conclusion section of the EAP teacher questionnaire.

Question	Rationale
12. How likely are you to use this test again? <i>very likely/likely/unlikely/very unlikely</i>	To discover if teachers thought it was a useful test, its re-use would suggest teachers thought it a useful assessment tool
13. Please indicate your general opinion of the test.	A general open question that allows attitudes to be given that may not have been given in elsewhere.

The questionnaire was designed so that most questions could be answered whether teachers had used the test or not.

3.12ii Questionnaire 2: EAP students

The second questionnaire was produced for EAP students. Its aim was to gauge their attitudes towards the test after having taken it. The format is similar to that for teachers which would enable subsequent comparable analysis between teacher and student attitudes towards the test.

The first part aims to elicit background information about the respondents.

Table 7. EAP student ‘introduction’ questions

Question	Rationale
1. Please indicate if you are: pre-sessional undergraduate / pre-sessional postgraduate (non PhD) / pre-sessional PhD / other (please specify)	To obtain information on level of study since the test was designed with postgraduate students in mind.
2. Please indicate your nationality:	To obtain demographic information about the respondents.
3. Please indicate the country in which you currently study:	
4. Please indicate your gender:	

In Part 2, students are asked their opinions on the test following the usefulness criteria proposed by Bachman & Palmer (2009). Although similar to the questions for teachers, the questions focus on the attitudes towards usefulness from the students’ perspective.

Table 8. EAP student ‘overall test evaluation’ questions

Question	Rationale
5.1 The test measures the types of speaking tasks that are relevant to my academic studies.	Aims to measure whether students believe the test has construct validity.
5.2 The content of the test is relevant to my academic studies.	This would reveal student attitudes towards the test’s authenticity.
5.3 The test does not require any special resources.	Aims to measure if the students see the test as practical.
5.4 I had the same chance as the other candidate to demonstrate my speaking ability.	Gauges student attitudes on test reliability.
5.5 If I did the test again, I would like to attend specific test preparation classes.	If specific classes were needed, this would suggest the test has a strong washback effect.
5.6 The test allowed me to demonstrate my academic speaking ability.	Indicates the level of interactiveness students have with the test.

Similar to the teachers, the study aimed to gain attitudes towards the questions with an emphasis on question 2. In particular, the use of countries as content and the false assertion. Hence the following questions were included.

Table 9. EAP student ‘Item evaluation’ questions

Question	Rationale
5. In question 1 of the test, you spoke to the other candidate in a more informal way. This is relevant to my academic speaking ability. (Strongly agree/agree/disagree/strongly disagree)	To gain student insight as to the relevance of using more informal language as part of an EAP programme.
6. Did you notice [the false assertion]? YES/NO	To gain data on how many students noticed the false assertion and compare this with teacher answers.
6.1 If yes, did you correct the examiner? YES/NO	To gain data on the number of students who challenged the false assertion.
6.2 The false assertion is helpful for my speaking abilities. <i>Strongly agree / agree / disagree / strongly disagree</i>	To gain student attitudes towards this feature of the test.

Finally, general opinions on the test which had not been given elsewhere were sought.

Students were also asked if they would like to sit the test again, which would give a general idea of the degree to which they enjoyed the test.

Table 10. Conclusion questions for EAP students

Question	Rationale
8. I would like to do the test again. <i>Strongly agree / agree / disagree / strongly disagree</i>	To reveal general attitudes about doing the test again.
9. Please indicate your general opinion of the test.	An open ended question to finish that allows students to state opinions they may not have had the chance to reveal elsewhere.

3.13 Validity and reliability

As Cohen et al. (2011: 179) put it “validity is an important key to effective research”.

Validity concerns the degree of accuracy of a research instrument; whether it actually measures what it sets out to measure. Dörnyei (2007: 48) calls validity the degree of credibility or truthfulness of a given piece of research.

Along with validity, reliability is an important concept and concerns the dependability of a measurement. In particular, whether similar measurements would be made at different times or places, and whether another researcher would make the same interpretation. Validity and

reliability are related concepts: reliability is seen as necessary for validity, whereas validity is not necessarily a precondition for reliability (Cohen et al. 2011: 179).

Validity and reliability are generally taken to be relative and not absolute. Threats to validity and reliability can seldom be removed completely but giving them attention throughout a piece of research can help mitigate them (Cohen et al. 2011: 179). In the context of this research, validity and reliability come into play for both the questionnaires and the test administration.

The main validity concerns for questionnaires are summed up by Cohen et al (2011: 209): whether respondents are giving truthful answers and whether they are representative of the target population.

Attempts were made in this research to address these questions. In the former question, respondents and their institutions were anonymous and they were clearly informed of this fact. Further, in the introduction to the questionnaires, respondents were encouraged to be as truthful as possible in their answers since this would aid the research (cf. questionnaire introduction Appendix 2) While these solutions are clearly not fool proof in ensuring truthful answers, it is hoped they would increase the possibility of gaining accurate answers.

The former question could be addressed by a follow up interview in which the researcher may press the respondent to clarify meaning and be more satisfied with a respondent's answer and understanding of the question. Regrettably, this research was unable to carry out follow up interviews due to time constraints. Yet it is acknowledged this would have enhanced the research's validity.

In addressing the second question, attempts were made to reach as wide a target population as possible by using the BALEAP mailing system. This has over 1000 subscribers from around the world and from a diverse range of teaching contexts.

In addition, steps recommended by Cohen et al. (2011: 209) were taken to gain a greater number of respondents: Follow up emails were sent to gently encourage participants, several emails were addressed personally so avoiding blanket general emails, the questionnaire was kept short and questions straightforward to answer.

3.14 Data analysis

Questionnaires were analysed using a simple descriptive method. For the quantitative closed questions this is necessary as less than 30 respondents were obtained which limits the use of

statistical analysis (Wallace 1998: 131). As a result, the answers given to the open questions were analysed for insightful comments given.

3.15 Ethical considerations

All research, particularly social research, concerns ethics since it is interested in people's lives in the social world (Dörnyei 2007: 63). Before contacting any participants an ethical approval form was completed and approved by the dissertation supervisor.

The key ethical issues were also highlighted in the introduction to the questionnaires. The voluntary nature of the questionnaire was made clear and it was indicated that the identity of both individuals and their institutions would be completely anonymous. Further, it was highlighted that data resulting from the respondents' answers would appear in my dissertation and potentially some other academic outputs (e.g. journal papers). Respondents were encouraged to contact myself, the dissertation supervisor or the Dean of Faculty should they have any concerns. They were also given the option to see the completed research.

Students who participated in the research were given a notebook and chocolate to say thank you. They also received feedback on their speaking via the mark sheet which was emailed to them.

4. Results and Discussion

4.1 Introduction

This chapter reviews the data gathered on attitudes towards the test based on two questionnaires, one for EAP teachers and one for EAP students. Both questionnaires aim to answer the two research questions:

1. What are EAP teachers' attitudes towards the test?
2. What are EAP students' attitudes towards the test?

A simple descriptive analysis is made for the responses from each questionnaire which contain a mixture of open and closed questions.

Reference is also made to the published literature and past studies on testing in EAP.

4.2 EAP teachers

A total of 10 EAP teachers completed the questionnaire. The respondents were a range of nationalities: 6 were British, 1 was Czech, 1 Hungarian, 1 German and 1 from the USA. Respondents were also asked where they currently worked. This revealed 7 worked in the UK, 1 respondent worked in Germany and 1 worked in the USA. One respondent chose not to answer this question.

Demographic information was obtained on the gender of respondents. This showed there were 5 female and 4 male respondents, with one declining to answer.

This part of the questionnaire also asked teachers whether they had used the test with their students. Unfortunately, only 1 respondent had actually used the test. This could be due to the timing of this research. The test and questionnaire were sent out in early July, a time when many EAP departments are in the midst of pre-sessional programmes. As a result, time would likely have been an issue for EAP teachers. Distributing the test at a different time of year may have yielded more test uses.

The one respondent who used the test indicated that it was administered to pre-sessional postgraduate students, the target student population for the test. While it would have been fruitful to interview this respondent, time constraints did not allow it.

This part showed that respondents were drawn from various EAP contexts and, as a result, would likely be working on different programmes within different institutions and in different countries. As highlighted in the literature, EAP is defined by being highly contextual (Hamp-Lyons 2001). Hence a variety of opinions would be expected due to respondents' differing teaching contexts.

4.3 EAP students

Sample information was also collected from student respondents. A total of 14 EAP students answered the questionnaire based on having done the test with myself as examiner.

Of these, 10 were from China, 2 were from Greece, 1 from Japan and 1 from Cyprus. While this may seem imbalanced, it represents the fact that China currently contributes the greatest number of international students to UK higher education (Universities UK, 2015).

Respondents were also asked the country in which they studied. All respondents answered either 'Scotland' or 'the UK' with the exception of one who answered with the name of the city in which she/he studied, rather than the country. While this could well have been a simple oversight, it potentially flags up a drawback of questionnaire use highlighted by Cohen et al. (2011: 383) and Dörnyei (2003: 11) that questionnaires' efficacy can be limited by the literacy of the respondents. This seems worth bearing in mind given that respondents are learners of English.

Respondents were also asked their current level of study. 9 indicated they were pre-sessional postgraduates. The remaining 5 selected the 'other' option. These were current postgraduate students on an in-sessional EAP course.

4.4 Overall test evaluation

The first part of the questionnaire aimed to obtain teachers' and students' attitudes towards test usefulness as defined by Bachman & Palmer (2009). This consisted of the 6 criteria previously defined: validity, authenticity, reliability, impact, interactiveness and practicality. Questions were carefully crafted to measure teachers' and students' opinions on each of these criteria.

Respondents were asked to indicate their level of agreement with 6 statements pertaining to the usefulness criteria, some of which are discussed below.

i. Validity

Teacher participants' responses to the first statement, which aimed to gauge attitudes towards the test's construct validity are shown in the figure below.

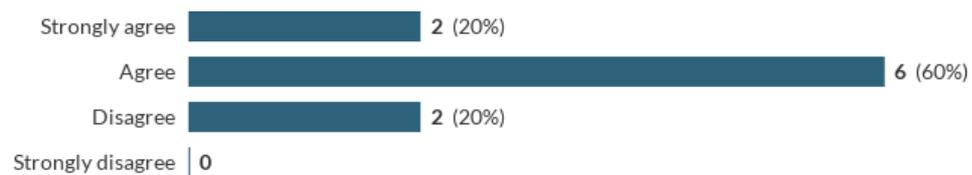


Fig 2. EAP teachers' attitudes towards the statement: *“The test measures the types of speaking tasks that EAP students will need in their academic studies”*.

As can be seen, a majority of respondents either agreed or strongly agreed with the statement. This indicates that among the respondents there was a belief in the validity of the test. This is encouraging, but also a feature of classroom-based tests; they are generally more valid and authentic compared to large scale proficiency tests which aim for greater reliability, often at the expense of validity. As highlighted in the literature review O’Sullivan (2011 cited in Schmidt & Hamp-Lyons 2015: 5) claims that local tests are more likely than proficiency exams to allow us to make valid assumptions about test takers.

Students were also asked their attitudes towards the test's validity in a similar question, albeit worded slightly differently.



Fig 3. Bar chart showing students' attitudes towards the statement *“The test measures the types of speaking tasks that are relevant to my academic studies”*.

This shows that, although the majority agreed with the statement, their feeling was more tepid than the teachers'. 5 student respondents indicated they thought the test does not measure academic speaking tasks, which although a minority is still relevant. However of these 5

respondents, 3 were pre-sessional students, so it could be argued that these students do not yet know what types of speaking tasks are expected of them at postgraduate level.

The general trend for the respondent sample of both teachers and students is agreement that the test's tasks are valid.

ii. Authenticity

Attitudes towards the authenticity of the test were sought from both students and teachers.

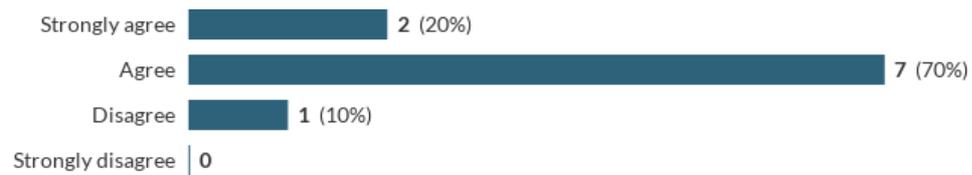


Fig 4. Bar chart showing EAP teacher attitudes towards the statement: “*The content of the test is relevant to EAP students*”.



Fig 5. Bar chart showing EAP student attitudes towards the statement: “*The content of the test is relevant to my academic studies*”.

This shows that the majority of teachers in the sample thought the test had authenticity, with 9 respondents either agreeing or strongly agreeing with the statement.

iii. Reliability

Along with authenticity and validity, reliability is also a key criteria of test usefulness. It is generally assumed that with increasing validity comes reduced reliability, and vice-versa (Bachman & Palmer 2009: 17). Hence, it would be interesting to discover perceptions on this given the general attitudes that the test is valid.



Fig 6. EAP teacher attitudes towards the statement: “Both candidates have the same opportunity to demonstrate their speaking ability”.



Fig 7. EAP student attitudes towards the statement: “I had the same chance as the other candidate to demonstrate my speaking ability”.

As the bar charts above show, the vast majority of both teachers and students either agreed or strongly agreed with the idea that the test was reliable in providing candidates with equal chances to demonstrate their speaking skill.

Test reliability has several facets, such as reliability among different raters, between different settings and between individual candidates. All of these contribute to a test’s overall reliability.

iv. Impact

Like reliability, impact has a number of facets. For example, impact on the individual, on the institution and on society (Bachman & Palmer 2009: 20). This question focused on the idea of ‘washback’, that is, whether learning and teaching would need to be modified to accommodate the test (Alexander et al. 2008: 308). This is often the most immediate impact for students and teachers.

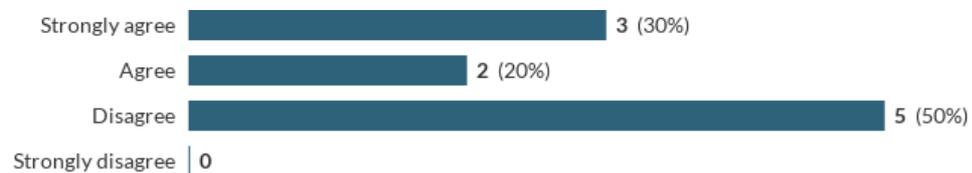


Fig 8. Teacher attitudes towards the statement: “If the test was mandatory, it would require specific test preparation classes”.



Fig 9. Student attitudes towards the statement: “If I did the test again, I would like to attend specific test preparation classes”.

As can be seen in the bar charts above, there is a lack of consensus between students and teachers on whether the test would require preparation classes. Preparation classes would imply the test content would have a washback effect and thus a strong impact on teachers and students. Both students and teachers were divided on the issue, perhaps reflecting different working or learning contexts.

As indicated in the sample description, students were a mixture of both pre-sessional and in-sessional students. It was hypothesised that the in-sessional students, given their relative experience of studying in English, would deem preparation classes less important than pre-sessional students, who may be newer to academic language. The data suggested this was generally the case: Of the 9 pre-sessional respondents, 7 agreed with the statement about needing preparation classes. Whereas of the 5 in-sessional students, only 1 agreed with the statement.

4.5 Evaluation of test items

The second part of the questionnaires aimed to gain greater depth of attitudes towards specific aspects of the test. It contained both closed and open questions which would allow some quantifiable data along with some more detailed insights.

As pointed out in the literature review, several authors (e.g. Jordan 1997, de Chazal 2014, Alexander et al. 2008) see informal language as important in EAP despite it not receiving much attention in traditional EAP courses. Hence this question aimed to obtain attitudes towards the inclusion of informal language in an EAP test.

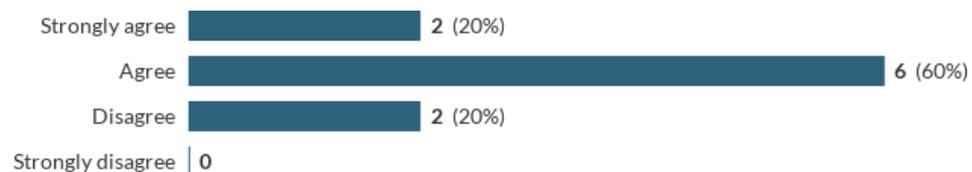


Fig 10. Bar chart showing teachers' attitudes towards the statement: “[Informal language] is appropriate for an EAP test of speaking”.



Fig. 11. Bar chart showing students' attitudes towards the statement: "[Informal language] *is relevant to my academic speaking ability*".

As the pie charts show there was a general belief that informal language is relevant to academic speaking ability. However, the wording in each question could negate a direct comparison. Nonetheless, respondents in this sample support the notion that more informal language is relevant in EAP and that it is an appropriate part of EAP speaking tests.

As indicated previously (cf. Test Design, Section 3.4), Question 2 of the test contained the inclusion of a false assertion by the examiner in a discussion, a feature about which teacher and student attitudes were sought.

8 student respondents claimed to have noticed the examiner's false assertion, while 5 indicated they did not notice it. These roughly balanced results are encouraging from a test design point of view, suggesting that the feature is not too difficult nor too easy.

As noted, only 1 teacher respondent used the test with students. However, this respondent indicated that the students did indeed notice the false assertion. Unfortunately, limited data nullifies any significance of teachers' use.

Both teachers and students were asked whether they thought this false assertion was relevant to an EAP test of speaking. Only 2 teachers answered this question, as shown in the figure below.



Fig 12. Bar chart showing teachers' attitudes towards the statement: "*This false assertion is appropriate for an EAP speaking test*".

Data from the teachers is fairly inconclusive given the low number of respondents, but also displays mixed feelings.

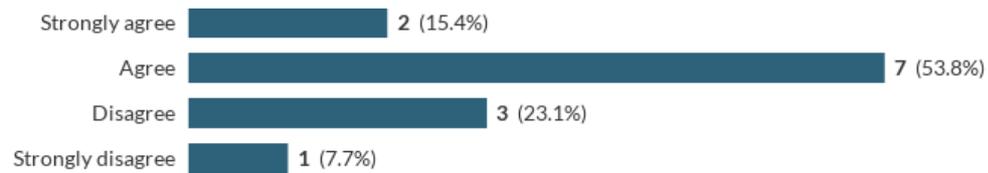


Fig 13. Bar chart showing students' attitudes towards the statement: "*The false assertion is helpful in developing my academic speaking abilities*".

The bar chart above shows mixed feeling among student respondents towards this feature, but with a bias towards believing the false assertion is a useful feature for academic speaking ability.

However, teachers were asked their opinion on the false assertion feature by way of an open question. Comments were mixed, but contained praise:

"Clever and creative". (Teacher respondent 9¹).

Slight confusion:

"I understand why it has been included (I think)...²" (Teacher respondent 3).

And reservations:

"I think this is problematic. It seems contrived to me and it doesn't really reflect an authentic situation". (Teacher respondent 10).

A frequent criticism was that students from particular cultures would find it strange to challenge the authoritative figure of the examiner. As illustrated in the following teacher comments:

"I worry that...Student's from countries where challenging others is less acceptable may be at a disadvantage here". (Teacher respondent 4).

"This will be a major issue for some students who hold teachers somewhat in reverence (SE Asian students especially)". (Teacher respondent 7).

While these observations are valid, it could also be argued that students should try and overcome these issues and adapt to the institution in which they are studying.

1

2 Teacher and student respondents were assigned numbers and letters, respectively.

Another claim was that if students were to sit the test a second time they would be primed and ready for the false assertion.

“However, once the candidates get used to the false assertion, I am not sure it will still provide same outcomes”. (Teacher respondent 9)

The rationale of the false assertion was to promote student listening skills in a discussion so as to integrate speaking and listening which de Chazal (2014) argues are “completely inseparable”. Hence if the test became a regular, occurrence examiners may choose to make a false assertion or not; but either way candidates would be listening attentively and critically in case it was included; the objective of this feature.

4.6 Countries as content

Teachers were asked their opinion on whether countries represented good content for the test.

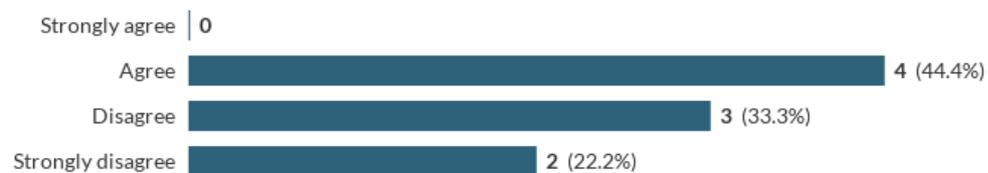


Fig 14. Teacher attitudes towards the statement: “*Countries are relevant content across students’ subject disciplines*”.

As the bar chart shows, the majority of teacher respondents thought that countries were not appropriate for all subject disciplines. They were asked an open question on their general opinions of countries as content in the test, some of which are illuminating:

One respondent saw countries as more suited to humanities/arts students, putting students from science disciplines at a disadvantage:

“Students specialising in pure/applied sciences, engineering or medicine, for example, will be disadvantaged. This tends to be overlooked by test developers, who, due to their disciplinary background, tend not to be comfortable with or knowledgeable about what would suit science/engineering/medical/etc. students.”

(Teacher respondent 8).

While an insightful comment, one would surely place geology, geography, zoology, tropical medicine and civil engineering, among others, as disciplines in which countries are extremely

relevant. Further, the test's country profiles contain various numerical data (e.g. population figures, GDP data and important years) and as Jordan (1997: 204) argues "verbalising data is important, particularly for those in the sciences". However, it is widely acknowledged that it is tricky to find texts general enough yet academic enough for EGAP purposes (Brindley & Ross 2001: 152).

Teacher respondent 10 believed that countries were not suitable content:

"I understand the reasons for choosing countries but I'm not sure it's especially academic."

Hence, for this respondent, countries do not conform with the view put forward by Alexander et al (2008: 243) that any topic used should be one "which can be discussed with academic rigour". However, the fact that other participants rate the content as valid suggests that more data need to be collected before a conclusion is reached.

Respondents were also asked for their attitudes towards the use of lesser known countries. The test also used lesser-known countries, Tanzania and Guyana. The thinking behind this was that it would galvanize an exploratory discussion as countries would likely be unknown.

As the bar chart below shows, teacher respondents generally agreed that lesser known countries facilitate discussion:

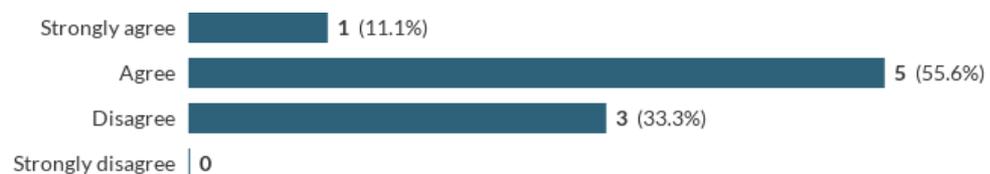


Fig 15. Bar chart showing teacher attitudes to the statement: "*Lesser known countries (i.e. Guyana and Tanzania) facilitate discussion*".

Several respondents made the point that 'lesser-known' countries is a relative term and students could well be from or near these countries:

"Stating the obvious, but these wouldn't be lesser known countries for students from Guyana and Tanzania. Wouldn't students from neighbouring African and South American countries be reasonably familiar with these countries?" (Teacher respondent 3).

"My only worry is that students who are from the country or neighbouring countries would have an unfair advantage". (Teacher respondent 4).

While what the teachers say is true, these two countries were used as an example. Relatively few international students from South America (1.2%) and Africa (6.1%) study in the UK (Universities UK 2015). However, other countries could easily be used to suit a particular class or group.

Several other teachers seemed to think that countries as content are suitable for EAP, with the idea perhaps not original as initially thought, as one respondent commented:

“Using countries as content is something we've long included as a possible task in our one-on-one diagnostic speaking test.” (Teacher respondent 9).

Another respondent seemed to display a positive attitude towards countries as content:

“The choice of the theme for a speaking test is possible and neutral enough to be administered with EAP students of difference Backgrounds).” (Teacher respondent 2).

As the comments show, there were differing opinions on countries' use as content. This could well reflect the different contexts the teachers work in. As one respondent flagged up it may not be appropriate for some students' subject discipline.

Several teachers may have worked with students from South American or East African backgrounds hence their comment on the 'lesser-known' status of the chosen countries. Nonetheless, other teachers seemed more positive on countries' utility as neutral and relevant enough to be used within EAP.

The diversity of opinion reflects the difficulty in finding suitably general topics with the necessary academic rigour for EGAP students. As the data indicate that a majority of teacher respondents believed countries were not relevant topic for EAP students

4.7 General attitudes

In the final part of each questionnaire, both teachers and students were asked an open question on their general attitudes towards the test. This would hopefully give respondents the chance to reveal attitudes towards the test that were not addressed elsewhere.

All respondents gave very useful comments. Among the themes to emerge were:

- The grading sheet needs improvement. In particular, it is currently lacking a section on pronunciation:

“I was missing an element of (academic) speech production that also contributes to intelligibility and comprehensibility, i.e. aspects/ descriptors of pronunciation at both segmental but mainly supersegmental level”. (Teacher respondent 10).

“I had some doubts about the criteria - shouldn't the range of structures used be measured and there's nothing about pronunciation here”. (Teacher respondent 2).

This is a good observation and while pronunciation could be considered under fluency (cf. marking criteria, Appendix 1), it is important enough to deserve a category of its own, particularly for a speaking test. As

- There was also some concern raised over the criteria used for Question 2, with one respondent arguing that presentation structure should not be considered a speaking skill *per se*:

“As it's a diagnostic test, is it fair to mark students on presentation structure - is that a speaking skill?” (Teacher respondent 10).

- Several respondents felt certain aspects involved elements that should not be important in a test of speaking or are not strictly academic. For example:

“2 parts (pre lect; and discuss work to do) while useful for students are parts that wouldn't be part of academic classroom discourse or assessment”. (Teacher respondent 6).

- Several teachers commented on the merit of the tasks but also made useful recommendations on ways in which they could be altered to achieve functions more relevant to EAP, for instance:

“...more focus on problem-solving or discussing in relation to input would be good”. (Teacher respondent 6).

“An important part of academic speaking is not just the description and presentaion of fact but how they can develop ideas and explore understadnings, develop and appropriately express a stance, propose, reject and amend propositions, plan a course of action in some detail etc. A lot of these points eg decide where to meet require little discussion "let's meet in the library at 6" "ok" or "Mt Everest is in Tanzania" (gosh my tutor's daft) response: mhm. This is not a discussion in eg seminars”. (Teacher respondent 7).

While problem solving and exploratory discussions were intended as part of the construct of the test, these respondents' comments highlight that perhaps these were not made explicit enough in the tasks. Modification of tasks could therefore help to bring out these parts of EAP speaking, such as clarifying the notes for examiners document.

Students were also given an open question to express their general attitudes towards the test. These comments did not contain as much constructive criticism as the teachers' comments. This is to be expected since teachers would likely have experience with testing and possibly test construction, whereas students are at the receiving end.

Examples of some generic positive comments included:

"I think that the test could be very efficient and helpful for pre-sessional students".
(Student respondent H).

And:

"It is helpful to prepare for the university study". (Student respondent A).

However, some light critique was also given concerning the instructions from the examiner:

"If possible, the interviewer could guide the test more clearly and encourage students to speak more". (Student respondent K).

This comment could perhaps reflect the need to better introduce candidates to the format of the test.

Overall, based on the data it could be suggested that student and teacher attitudes to the test were positive. Teachers seemed to think the test had potential but some items and procedures would need reworking:

"This is an interesting start and has a lot of potential". (Teacher respondent 2).

"has good potential". (Teacher respondent 6).

"It is a good attempt to make the testing of speaking more relevant to EAP students".
(Teacher respondent 8).

However, the current shortfalls in the test might dissuade teachers from using the test again, as the data suggest:



Fig. 16. Teacher attitudes to the question: “*How likely are you to use the test again?*”

However, this should be compared with the overwhelmingly positive student feedback about their attitudes towards doing the test again:

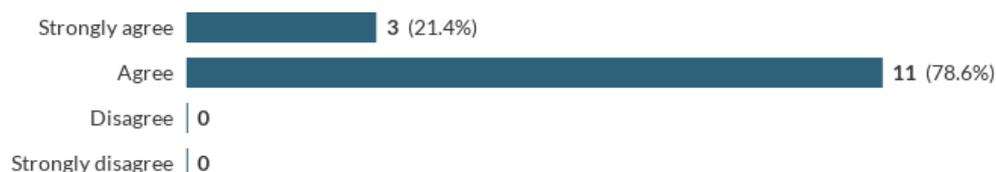


Fig. 17. Student attitudes towards the statement: *I would like to do the test again.*

The analysis revealed teachers’ and students’ attitudes towards the test that could contribute to its further development. It is worth noting that I personally administered the test to the EAP students. This could have been a factor in their responses; perhaps they wanted to be complementary. The teachers were anonymous and therefore had no such issues. Additionally, teachers in general did not use the test with their students, so their answers could have been different had they used the test in the classroom.

As indicated, the teachers’ differing attitudes on issues such as countries being relevant content could be due to the different contexts in which they work or have worked. Countries may well be suitable for some EAP contexts, but perhaps not all.

Teachers also indicated their generally negative attitudes towards the false assertion feature of the test. This would suggest the item needs rethinking.

Teachers also made valuable comments regarding the marking system. In particular, more guidance on criteria and an opportunity to mark pronunciation were cited as both relevant and desirable.

5. Conclusion

This research has looked at EAP teacher and student attitudes towards a self-designed diagnostic test of speaking for EAP. It was hoped that the test could be useful to some teachers in certain contexts and contribute to debates on testing speaking in EAP more generally.

The research was guided by two questions, namely:

1. What are EAP teachers' attitudes towards the test?
2. What are EAP students' attitudes towards the test?

Two questionnaires, one for the students and another for the teachers, asked a range of questions that gauged their respective opinions on the test.

In general, it can be stated that teachers gave positive feedback on the test and in seeing new tests of this kind being developed (cf. Section 4.7). Many were galvanised to write a considerable amount about the test. This arguably shows that the test seems to have fulfilled one intended function: to contribute to the wider debate about testing in EAP.

A key component of the questionnaires was questions concerning the six criteria of usefulness as proposed by Bachman and Palmer (2009). These concern the validity, reliability, impact, interactiveness, practicality and authenticity of any test. Questions for both students and teachers were as similar as possible to enable a comparison of answers given.

For some criteria, there was a clear consensus in opinion among respondents: the vast majority of teachers (10) and students (12) believed the test had reliability. However, while 9 teachers thought the test was authentic, only 8 students thought the test fulfilled this criterion. These data suggest a divergence of opinion, possibly due to pre-sessional students not having a good handle on what exactly constitutes academic content. Similarly, there was dissonance within teacher and student respondents about the test's impact: about half of all student respondents and half of all teacher respondents believed the test had a strong impact. This division would also be worth exploring further in order to discover the reasons for the differing attitudes. Follow-up interviews would perhaps have enabled a fuller explanation of this.

The questionnaire data also showed that, on the whole, students enjoyed the test. This is an important factor, and it is tentatively suggested that ‘tests’ may not always equate to anxiety among students. Several features of the test, such as the informal start and personable approach of the examiner, perhaps contributed to this.

Despite the positivity, much constructive criticism was also given in relation to specific aspects of the test. The false assertion feature received, on the whole, negative feedback from EAP teachers and a lukewarm response from students. Although the idea for the item was not dismissed entirely, the data suggest it would need considerable reworking in order to become a more valid and reliable test feature.

Other sensible concerns were raised over the grading criteria, such as the fact pronunciation was not made explicit and that some criteria (e.g. presentation structure) were not strictly part of speaking.

The study suffered from a number of limitations. Firstly, the relatively small sample limits the extent to which any findings can be generalised. Another limitation was the limited uptake of the test by EAP teachers. While 10 teachers completed the questionnaire, only 1 did so having actually used the test. More test users would have resulted in more in-depth data. It is also relevant that among the 14 student respondents 10 were Chinese, so student questionnaire responses are dominated by this one nationality. Yet while this is a limitation, there is a very large Chinese student cohort in the UK (Universities UK 2015). The mixture of both pre-sessional and in-sessional students could also be considered a drawback. These two groups are at different stages of their learning and greater coherence in results could have been gained by focusing on just one group. As mentioned, timing was a key issue. A longer period (e.g. an entire academic year) might well have yielded more interesting data as teachers from a variety of contexts would have had the possibility to use the test and give their feedback. A considerable amount of time was spent on the initial test design and piloting. By the time it was ready to be shared, it was already peak time for EAP practitioners. This could explain the limited uptake.

However, the study had a strength in that there are currently few tests of speaking for EAP in circulation. As evidenced by several teachers’ enthusiastic responses (cf. Section 4.7), it is a topic about which many in EAP are concerned and interested. Making the test freely available online, therefore, hopefully made some sort of contribution. This also gives something in return for those who kindly volunteered their time to participate in the study.

Going forward, there are modifications that could be made to the test design based on users' (both student and teacher) comments. This would contribute to enhancing its quality and usefulness, and perhaps its uptake in the classroom. However, further pilot studies, test design work, user feedback and follow-up interviews would be needed before anything a close to a 'good test' was produced.

The research also raises a number of questions, suggested by the teachers' comments (cf. Section 4):

- Can a classroom/formative test be used across EAP contexts?
- Would a pool of test items be helpful for EAP teachers?
- What is suitable content for EGAP tests?
- Should informal language be part of EAP courses/tests?

In conclusion, the test has gone through a reasonably rigorous road of development and evaluation. Yet there still remains a lot of work to be done before the test could be considered useful, at least in most EAP contexts. This reflects the difficult journey test developers are on, with the constant balancing of criteria such as reliability and authenticity one of many pitfalls to be wary of. Nevertheless, the journey is a stimulating one that cuts to the heart of the matter of teaching: what should students learn and whether they have learned it.

References

- Alexander, O., Argent, S. & Spencer, J. (2008). *EAP essentials: A teacher's guide to principles and practice*. London: Garnet Education.
- Bachman, L.F. (1991). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F., & Palmer, A.S. (2009). *Language testing in practice*. Oxford: Oxford University Press.
- BALEAP. (2008). *Competency framework for teachers of English for academic purposes*. Available at: <https://www.baleap.org/wp-content/uploads/2016/04/teap-competency-framework.pdf>. [Accessed 14/05/2016].
- Brindley, G., & Ross, S., (2001). *EAP Assessment: Issues, models and outcomes*, pp148-166. In Flowerdew, J., & Peacock, M. (Eds). (2001). *Research perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press.
- Brindley, G. *Assessment*. In Carter, R. & Nunan, D. (Eds). (2001). *Teaching English to Speakers of Other Languages*. Cambridge: Cambridge University Press.
- Bristol Online Surveys. 2016. *Bristol Online Surveys*. [ONLINE] Available at: <https://admin.onlinesurveys.ac.uk/>. [Accessed 19 July 2016].
- British Council. (2013). *Infographic: student mobility*. Available at <http://institutions.ukcisa.org.uk/Info-for-universities-colleges--schools/Policy-research--statistics/Research--statistics/Infographic-on-student-mobility-trends/> [Accessed 06/05/2106]
- Cohen, L., Manion, L & Morrison, K. (2011). *Research methods in education*. London: Routledge.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cross, J. & Hitchcock, R. (2007). *Chinese students' (or students from China's) views of UK HE: differences, difficulties and benefits, and suggestions for facilitating transition*. The East Asian learner, Vol. 3, No.2.

Davidson, F., & Cho, Y. (2001). *Issues in EAP test development: What one institution and its history tells us*. In Flowerdew, J., & Peacock, M. (Eds). (2001). *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press.

Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell.

Dearden, J. (2014). *English as a medium of instruction – A growing phenomenon*. British Council. Available at <http://www.education.ox.ac.uk/wordpress/wp-content/uploads/2014/09/EMI-a-Growing-Global-Phenomenon-new-cover.pdf>. [Accessed 14/05/2016].

De Chazal, E. (2014). *English for academic purposes*. Oxford: Oxford University Press.

Doe, C. (2015). *One teacher's take on using a test for diagnostic purposes in the classroom*. *Journal of English for Academic Purposes*, Vol.18, June 2015, pp40-50.

Dörnyei, Z. (2003). *Questionnaires in second language research: construction, administration and processing*. London: Lawrence Erlbaum Associates.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Douglas, D. (2000) *Assessing language for specific purposes*. Cambridge: Cambridge University Press.

Dudley-Evans, T., & St John, M.J. (1998). *Developments in English for specific purposes*. Cambridge: Cambridge University Press.

Fox, J.D. (2009). *Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment*. *Journal of English for Academic Purposes*. Vol 8 (1) 28-42.

Hamp-Lyons, L. (2001) *English for academic purposes*. In Carter, R. & Nunan, D. *Teaching English to Speakers of Other Languages*, pp126-131. Cambridge: Cambridge University Press.

Heaton, J.B. (1994). *Writing English language tests*. Harlow: Longman.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hutchinson, T., & Waters, A. (1996). *English for specific purposes*. Cambridge: Cambridge University Press.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London: Taylor & Francis.

Jameson, J., Wang, L. & Church, J. (2013). *In-house or commercial speaking tests: Evaluating strengths for EAP placement*. *Journal of English for Academic Purposes*. Vol 12 (4), December 2012, pp288-298.

Jordan, R.R (1997). *English for academic purpose: A guide and resource book for teachers*. Cambridge: Cambridge University Press.

Kennedy, C. & Boltho, R. (1996). *English for specific purposes*. Basingstoke: Macmillan.

Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge: Cambridge University Press.

McNamara, T. (2011). *Language testing*. Oxford: Oxford University Press.

Moss, P. (2003). *Reconceptualizing validity for classroom assessment*. *Educational measurement: Issues & Practice*. Volume 22, Number 4, pp13-25.

Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.

Ockey, G.J., Koyama, D., Setoguchi, E. & Sun, A. (2015). *The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students*. *Language Testing* 2015, Vol. 32(1) 39–62.

Robinson, P., Strong, G., Whittle, J. & Nobe, S. (2001) *The development of EAP oral discussion ability*. In Flowerdew, J., & Peacock, M. (2001). *Research Perspectives on English for Academic Purposes*, pp347-359. Cambridge: Cambridge University Press.

Schmitt, D., & Hamp-Lyons, L. (2015). *The need for teacher knowledge in assessment*. *Journal of English for Academic Purposes*. Vol 18, June 2015, pp3-8.

Swales, J.M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Underhill, N. (2004). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.

UNESCO (2016). *Global flow of tertiary level students*. Available at: <http://www.uis.unesco.org/Education/Pages/international-student-flow-viz.aspx>. [Accessed 17/05/2016].

Universities UK (2015) *Higher education in facts and figures 2015* Available at: <http://www.universitiesuk.ac.uk/facts-and-stats/data-and-analysis/Pages/higher-education-facts-and-figures-2015.aspx>. [Last accessed 10/08/2016].

Wallace, M.J. (1998). *Action research for language teachers*. Cambridge: Cambridge University Press.

Weir, C.J (2005). *Language testing and validation: An evidence based approach*. London: Palgrave Macmillan.

Appendix 1: Test pack (Test paper, country profiles, notes for examiner and mark sheet)

Student copy (Candidate A)

English for Academic Purposes (EAP) Diagnostic Test of Speaking

*This test is designed to help you in your future studies at university. An important part of success in academic study is clarifying information and asking questions if you do not understand something. These factors are graded in this test. So please **ask the examiner any questions** about the test **at any time**. You will receive feedback from the test.*

*Please remember this is a **language** test. You are **not** assessed on your content knowledge.*

Question 1

Turn to your partner.

Imagine you are both in the same class and are waiting for a lecture to begin. You do not know each other. Introduce yourself and get as much information as possible about your classmate.

You have about 3 minutes for this. Try to spend about **half the time** talking about you and the other half talking about your partner.

Possible topics include:

- Hometown/country of origin.
- Current studies.
- Today's weather.
- Hobbies and interests.
- Homework for the class.

You have 15 seconds to prepare and then 3 minutes to talk.

(3 minutes)

Question 2

You were given some information about a country to read. You will now each make a short 2 minute presentation to the other candidate and examiner about this country. You should include any information you think is interesting or relevant. Try to structure your talk in a clear way, with an introduction, main part and a conclusion.

Part A

Candidate A: It is your turn to speak first. After your talk you will have a short discussion about the country with the other candidate. Then you will both have a short discussion with the examiner.

You may ask the examiner any questions about the text now.

Candidate A: You may start your talk now. You have about 2 minutes.

(2 minutes)

Candidate A: The other candidate will now ask you questions about your talk.

(2 minutes)

You will both now have a short discussion with the examiner about the country.

(3 minutes)

Part B

Candidate A: You are going to listen to candidate B talk about a country. You must listen carefully and you may take notes. At the end of the talk you must start a discussion about the country. This could include asking a question, clarifying or checking any information, giving opinions, etc.

(2 minutes)

Candidate A: You may now ask Candidate B any questions about the talk in order to begin a discussion. You have about 2 minutes.

(2 minutes)

You will both now have a short discussion with the examiner about the country.

(3 minutes)

Question 3

Now you are going to talk to your partner again. You are in the same class and must work together on an upcoming group assignment. Your assignment is below. Read this carefully.

Assignment:

You will make a group presentation next month about the works of Charles Dickens, one of the English language's most famous authors. Your presentation must include a detailed biography of Dickens, an abridged bibliography, major themes that emerge within his oeuvre, your own critique (such as your favourite or least favourite novels) and some relevant pictures. You must use PowerPoint to deliver your presentation.

(1 minute)

You may ask the examiner any questions about the assignment now.

Now, you are going to talk to your partner about the assignment. You must:

- Decide who will do each task.
- Talk about where you might find the relevant information.
- Agree on a suitable time and place to meet and discuss your progress.

You may start talking to your partner now.

(2 minutes)

-----End of Test-----

Student copy (Candidate B)

English for Academic Purposes (EAP) Diagnostic Test of Speaking

*This test is designed to help you in your future studies at university. An important part of your success in academic study is clarifying information and asking questions if you do not understand something. These factors are graded in this test. So please **ask the examiner any questions** about the test **at any time**. You will receive feedback from the test.*

*Please remember this is a **language** test. You are **not** assessed on your content knowledge.*

Question 1

Turn to your partner.

Imagine you are both in the same class and are waiting for a lecture to begin. You do not know each other. Introduce yourself and get as much information as possible about your classmate.

You have about 3 minutes for this. Try to spend about **half the time** talking about you and the other half talking about your partner.

Possible topics include:

- Hometown/country of origin.
- Current studies.
- Today's weather.
- Hobbies and interests.
- Homework for the class.

You have 15 seconds to prepare and then 3 minutes to talk.

(3 minutes)

Question 2

You were given some information about a country to read. You will now each make a short 2-minute presentation to the other candidate and examiner about this country. You should include any information you think is interesting or relevant. Try to structure your talk in a clear way, with an introduction, main part and a conclusion.

Part A

Candidate B: You are going to listen to candidate A talk about a country. You must listen carefully and you may take notes. At the end of candidate A's talk you must start a discussion about the country. This could include asking a question, clarifying or checking any information, giving opinions, etc.

(2 minutes)

Candidate B: You may now ask Candidate A any questions about the talk in order to begin a discussion. You have about 2 minutes.

(2 minutes)

You will both now have a short discussion with the examiner about the country.

(3 minutes)

Part B

Candidate B: Now it is your turn to speak. After your talk you will have a short discussion with the other candidate. Then you will both have a short discussion with the examiner.

You may ask the examiner any questions about the text now.

Candidate B: You may start your talk now. You have about 2 minutes.

(2 minutes)

Candidate B: The other candidate will now ask you questions about your talk.

(2 minutes)

You will both now have a discussion with the examiner about the country.

(3 minutes)

Question 3

Now you are going to talk to your partner again. You are in the same class and must work together on an upcoming group assignment. Your assignment is below. Read this carefully.

Assignment:

You will make a group presentation next month about the works of Charles Dickens, one of the English language's most famous authors. Your presentation must include a detailed biography of Dickens, an abridged bibliography, major themes that emerge within his oeuvre, your own critique (such as your favourite or least favourite novels) and some relevant pictures. You must use PowerPoint to deliver your presentation.

(1 minute)

You may ask the examiner any questions about the assignment now.

Now, you are going to talk to your partner about the assignment. You must:

- Decide who will do each task.
- Talk about where you might find the relevant information.
- Agree on a suitable time and place to meet and discuss your progress.

You may start talking to your partner now.

(2 minutes)

-----End of Test-----

Country Profile A



Country: Guyana

Population: 735, 555

Capital: Georgetown

GDP: \$6.15 billion (\$7,938 per capita)

Main industries: Agriculture (production of rice and sugar), bauxite mining, gold mining, timber and shrimp fishing.

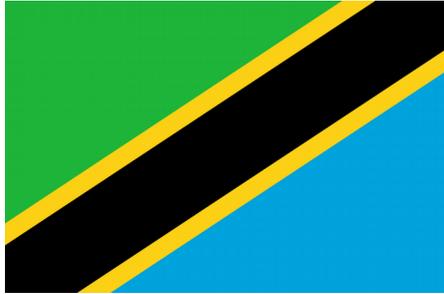
Guyana is situated in the northern part of South America and is the only nation on that continent where English is an official language. However, most people speak an English creole.

The country was home to several native tribes before being colonised by the Dutch in 1616. It remained under Dutch control until the British took over in 1796. The country then became a plantation economy, that is, the economy relied on exporting a handful of crops. Many slave workers were brought here from various parts of the British Empire. Guyana's present day population reflects this and is made up of people from Indian and African backgrounds, along with the indigenous people. It became independent from Britain in 1966 but remains part of the Commonwealth.

Guyana is considered part of the West Indies and often comes under its aegis, for example in international cricket, one of the country's most popular sports.

(Adapted from Wikipedia.org)

Country Profile B



Country: Tanzania

Population: 50.76 million

Capital: Dodoma

GDP: \$127.69 billion (per capita \$2,591)

Main industries: Mining (Natural gas, gold and diamonds), agriculture and tourism.

Tanzania is a country in eastern Africa. It is one of the world's poorest nations; hunger and malnutrition are commonplace, especially in rural regions.

However, Tanzania is rich in natural resources: multi-national companies mine gold, gas and diamonds and export them to the world market. Agriculture is also an important part of the economy. Maize, cassava and sweet potatoes are the largest crops and all are both exported and consumed domestically.

Tanzania was home to many indigenous people before being colonised: First by the Portuguese, then the Arabs, later the Germans and finally Britain, which took control after the First World War. The country became independent from Britain in 1961.

Tanzania has become popular with tourists visiting Mount Kilimanjaro, Africa's highest mountain. Despite its capital being Dodoma, the city of Dar es Salaam is better known. It houses most of the government and official buildings, and has the largest population in the country.

(Adapted from Wikipedia.org)

English for Academic Purposes (EAP) Diagnostic Test of Speaking

Instructions and brief rationale for examiner

(Please read in conjunction with the test, country profiles and mark sheet)

Introduction

This test of speaking has been designed for students about to embark on, or in the middle of, a course in English for Academic Purposes (EAP). It has been designed with postgraduate students in mind, but could also be used for undergraduates.

As a test of speaking, it is imperative that the examiner makes the candidates feel as relaxed as possible. This means adopting a friendly manner, wording any questions in a clear way and creating a relaxed atmosphere in which candidates are given the best chance possible to demonstrate their academic speaking ability.

Studies have also shown that examiners who take excessive notes during a test can affect candidates. Therefore, examiners are advised to take as few notes as possible. The mark sheet has been designed with this in mind.

The mark sheet offers a simple way for examiners to grade candidates' test performance. However, this is a classroom diagnostic test and not a formal proficiency exam. As a result, users of the test should feel free to modify timings, task delivery and grading as they see fit. The test aims to identify areas in which students could improve their academic speaking.

The test is designed to give a snapshot of student speaking ability. As such no specific preparation is required. But it is recommended to **give students the relevant question paper 24 hours before the test administration**. This reduces any anxiety students may have about the test.

Material needed:

Printed copies of:

Student copies of test (candidate A and candidate B versions)

Country profiles

Mark sheet

Notes for interlocutor (if required)

Pens

Blank paper

Instructions – Start of test

Welcome students and invite them to sit down. Ask them their names, nationalities, how they are and what they are doing today.

Randomly assign candidates as A and B. Hand students a **country profile**: give candidate A the Guyana country profile and hand candidate B the Tanzania country profile. Tell students they have 10 minutes to read this in private. Then leave the room or send students away to read it, as appropriate.

(10 minutes later)

Students return and the test begins. Issue students a copy of the test if they don't have one, **remembering to give candidate A and candidate B the relevant study copy**. This is indicated at the top of the page.

Read students the following aloud, which also appears on their copy of the test:

*This test is designed to help you in your future studies at university. An important part of success in academic study is clarifying information and asking questions if you do not understand something. These factors are graded in this test. So please **ask the examiner any questions** about the test **at any time**. You will receive feedback from the test.*

*Please remember this is a **language** test. You are **not** assessed on your content knowledge.*

Ask students if they have any questions and emphasise that they should not hesitate to ask questions at any point.

Question 1

Rationale

This is a relatively straightforward situation that most students should be able to complete without too much difficulty. It enables students to show their ability from the start, helping them relax into the test situation. Examiners should be listening for turn-taking strategies, evidence of understanding and students' ability to ask questions, as per the mark sheet.

Instructions

Direct students' attention to Part 1. Read the question aloud:

Turn to your partner.

Imagine you are both in the same class and are waiting for a lecture to begin. You do not know each other. Introduce yourself and get as much information as possible about your classmate.

*You have about 3 minutes for this. Try to spend about **half the time** talking about you and the other half talking about your partner.*

They have 15 seconds to think about their response and then 3 minutes to speak. Control the time.

Make appropriate insertions on the grading sheet, but aim to spend most of the time listening to the candidates rather than taking notes.

Question 2

Rationale

Students will have read the country profile (a short text giving information about a country) you gave them prior to the exam. They were given 10 minutes for this.

This part of the test asks candidates to deliver a 2-minute talk about what they have read. Examiners should listen for presentation structure along with empathy with audience; terms that are likely to be new to the other candidate should be explained, for instance.

Lesser known countries have deliberately been chosen as texts for two main reasons. Firstly, candidates' content knowledge of these countries may be lower than for other countries, which could stimulate questions and an exploratory discussion about the countries. Secondly, countries provide good content for EAP since most subject disciplines talk about countries in some way. Country profiles deliberately contain less common words that are relevant but perhaps are unknown to students. Again this provides greater stimulus for candidates to ask questions - a feature which this test aims to both encourage and test.

Each candidate gives a 2 minute talk about the country. Then engages in a 2 minute discussion with the other candidate about the country. Finally, both candidates have a discussion with the examiner about the country.

In this discussion with the examiner, some example questions for each country are given below. This part of the test also contains a novel feature: **in your discussion with the candidates you will make a false assertion about the country in question.** This will be a factually incorrect statement based on what was said in the talk. Some example false assertions are given below. Aim to deliver the false assertion twice and as factually as possible. The idea is to check if students are paying attention and willing to challenge the examiner. If corrected, apologise for the mistake and move on.

In Part A, candidate A gives a talk first. Then in part B the process is repeated with candidate B's country.

Instructions

Instructions should be read as follows:

You were given some information about a country to read. You will now each make a short 2 minute presentation to the other candidate and examiner about this country. You should include any information you think is interesting or relevant. Try to structure your talk in a clear way, with an introduction, main part and a conclusion.

Part A.

Candidate A: *It is your turn to speak first. After your talk you will have a short discussion with the other candidate. Then you will both have a short discussion with the examiner.*

You may ask the examiner any questions about the text now. (**Answer questions as appropriate**).

Candidate B: *You are going to listen to candidate A talk about a country. You must listen carefully and you may take notes. At the end of the talk you must start a discussion about the country. This could include asking a question, clarifying or checking any information, giving opinions, etc.*

Candidate A: *You may start your talk now.*

(2 minutes) Control the time.

After two minutes: *Thank you candidate A.*

Candidate A: *The other candidate will now ask you questions about your talk.*

Candidate B: *You may now ask Candidate A any questions about the talk in order to begin a discussion. You have about 2 minutes.*

(2 minutes) Control the time.

After two minutes: *You will both now have a short discussion with the examiner about the country.*

(3 minutes)

Note for examiner: You MUST include a false assertion during this discussion. Examples of discussion questions and false assertions are included below.

Country: Guyana

Example discussion questions:

Have you ever heard of Guyana before?

What kind of jobs do you think people do in Guyana?

Would you like to visit the country? Why/why not?

Why do you think tourism isn't popular there?

Example false assertions. Choose one. Or change a figure/date/name/commodity as appropriate:

I am surprised that silver is mined there.

The country only became independent in 1980. That seems so recent.

It's interesting that so many Indonesians live there.

Part B.

Candidate B: *Now it is your turn to speak. After your talk you will have a short discussion with the other candidate. Then you will both have a short discussion with the examiner.*

You may ask the examiner any questions about the text now. (Answer as appropriate).

Candidate A: *You are going to listen to candidate B talk about a country. You must listen carefully and you may take notes. At the end of the talk you must start a discussion about the country. This could include asking a question, clarifying or checking any information, giving opinions, etc.*

(2 minutes)

Candidate B: *You may start your talk now. You have about 2 minutes.*

(2 minutes)

After two minutes: *Thank you candidate B.*

Candidate A: *You may now ask Candidate B any questions about the talk in order to begin a discussion. You have about 2 minutes.*

(2 minutes)

After two minutes: *You will both now have a discussion with the examiner about the country.*

Note for examiner: You MUST include a false assertion during this discussion. Examples of discussion questions and false assertions are included below:

Country: Tanzania

Example discussion questions:

Have you ever heard of Tanzania? Where/how?

Would you like to visit the country? Why/why not?

What were its main industries?

How does the country's GDP/industries/population compare to your own country?

Example false assertion. Choose one. Or change a figure/date/name as appropriate:

People often do not realise Mount Everest is in Tanzania.

It's a big country, with 200 million people it must have one of the largest populations in Africa.

I didn't know the capital was Dar es Salaam. Have you heard of that city?

Question 3

Rationale

This last part involves the two students working together. The situation is a group project that has about 5 or 6 tasks to complete. Students must discuss the assignment but also demonstrate their problem solving ability by suggesting relevant locations to find the different information and delegate the tasks between them.

The examiner should pay close attention to students' turn taking, problem solving and discourse management. The task requires the students to discuss 3 concrete points (i.e. "Decide who will do each task, talk about where you might find the relevant information and agree on a suitable time and place to meet and discuss your progress"). Examiners should note whether students address all of these points.

Instructions

Read the instructions to the candidates:

Now you are going to talk to your partner again. You are in the same class and must work together on an upcoming group assignment. Your assignment is below. Read this carefully.

Let them read the assignment on their own. Allow about a minute for this.

Ask them if they would like to clarify anything about the assignment.

Then read them the task aloud:

Now, you must talk to your partner about the assignment. You must:

-Decide who will do each task.

-Talk about where you might find the relevant information.

-Agree on a suitable time and place to meet and discuss your progress.

As in question 1, aim to listen attentively to candidates and avoid excessive note taking.

End of test: Thank candidates for their participation and allow a brief period for any questions or comments from them.

Mark Sheet (Candidate A)

Part 1

Vocabulary	0	1	2	3	4
Fluency	0	1	2	3	4
Accuracy	0	1	2	3	4
Turn taking	0	1	2	3	4
Evidence of understanding	0	1	2	3	4
Asking questions	0	1	2	3	4

Part 2

Vocabulary	0	1	2	3	4
Fluency	0	1	2	3	4
Accuracy	0	1	2	3	4
Presentation structure	0	1	2	3	4
Empathy with audience	0	1	2	3	4
Challenging examiner	0	1	2	3	4

Part 3

Vocabulary	0	1	2	3	4
Fluency	0	1	2	3	4
Accuracy	0	1	2	3	4
Problem solving	0	1	2	3	4
Turn taking	0	1	2	3	4
Discourse management	0	1	2	3	4

Briefly indicate your overall impression of the candidate's speaking ability, specifically referring to any areas that could be improved:

Mark Sheet (Candidate B)

Part 1

Vocabulary	0	1	2	3	4
Fluency	0	1	2	3	4
Accuracy	0	1	2	3	4
Turn taking	0	1	2	3	4
Evidence of understanding	0	1	2	3	4
Asking questions	0	1	2	3	4

Part 2

Vocabulary	0	1	2	3	4
Fluency	0	1	2	3	4
Accuracy	0	1	2	3	4
Presentation structure	0	1	2	3	4
Empathy with audience	0	1	2	3	4
Challenging examiner	0	1	2	3	4

Part 3

Vocabulary	0	1	2	3	4
Fluency	0	1	2	3	4
Accuracy	0	1	2	3	4
Problem solving	0	1	2	3	4
Turn taking	0	1	2	3	4
Discourse management	0	1	2	3	4

Briefly indicate your overall impression of the candidate's speaking ability, specifically referring to any areas that could be improved:

Marking criteria

Examiners should mark using an ‘impression marking’ system; allocating a grade for each component on a scale of 0 – 4, where 0 indicates insufficient evidence displayed on which to judge the candidate and 4 shows a masterful display, based on subjective judgement.

An elaboration of each component is given below. At the end of the test, markers should note candidates’ major weak points, bearing in mind the diagnostic purposes of the test. This should then be given as feedback to candidates either verbally or in writing.

Vocabulary. Refers to a candidate’s range of vocabulary in a traditional sense; their ability to avoid repetition of the same words and evidence of advanced vocabulary. In particular, higher marks should be awarded for adopting an appropriate register for the context; more colloquial in question 1 and more formal in question 2.

Fluency. Candidates’ meaning should be clear and delivered in a confident fluid style. Hesitations should be part of speech rather than interrupt it excessively.

Accuracy. While the clarity of their message is more important, candidates should display a good knowledge of structure, using appropriate tenses and conditionals, for instance.

Turn taking. Candidates should show appropriate strategies for promoting a dialogue. These include asking questions, pausing and responding to what the other candidate said. Candidates should also show initiative in maintaining the dialogue by making relevant contributions.

Evidence of understanding. In the dialogue, candidates should display unambiguous evidence that they have understood and processed what the other candidate/examiner has said. A simple example of this: A: “*I really enjoy playing cricket*”. B: “*That’s interesting. When did you start playing?*”

Presentation structure. Although short, the candidates’ talk should include an appropriate introduction and conclusion. It should also contain appropriate signposting language. Candidates should also avoid excessive reading from the text, although they may refer to it.

Empathy with audience. Candidate shows both linguistic control and audience awareness. Since the audience is non-specialist any technical or unusual words or concepts should be explained.

Challenging examiner. Maximum marks would be awarded to a candidate who spots the false assertion and corrects the examiner in a clear and confident way.

Problem solving. Candidates should make plausible attempts to address the 3 tasks relating to the group assignment.

Asking questions. Candidates are able to ask questions accurately and for a variety of purposes (e.g. clarification, checking, interpretation check, digression, challenge).

Discourse management. Candidates’ ability to control the direction of the conversation.

	Band	0	1	2	3	4	
Criteria							
Vocabulary		Candidate does not speak enough to give a grade.	Candidate uses only basic vocabulary below the required level.	Candidate displays an adequate range of vocabulary but some improvement required.	Candidate displays a very good range of vocabulary.	Candidate displays very advanced vocabulary for the level.	
Fluency		Candidate does not speak enough to give a grade.	Candidate takes very long pauses and hesitations that detract from the message.	Candidate speaks such that the message is reasonably clear although more practice needed.	Candidate shows very good level of fluency.	Candidate is highly fluent in speech for the level.	
Accuracy		Candidate does not speak enough to give a grade.	Frequent grammatical errors that impede meaning.	Some grammatical errors but message not generally impeded.	Very good command of grammar across a range of structures.	Highly advanced knowledge displayed over a wide range of structures.	
Turn taking		Candidate does not speak enough to give a grade.	Candidate shows very little awareness of turn taking strategies.	Candidate shows some awareness of turn taking strategies.	Very good awareness of turn taking strategies.	Acute awareness of turn taking strategies.	
Evidence of understanding		No evidence of understanding shown.	Candidate seldom asks follow up questions or acknowledges other candidate's utterances.	Candidate sometimes asks relevant follow up questions.	Clear understanding shown by follow ups, relevant comments and referring back to prior comments.	Unambiguous evidence that candidate understands everything.	
Presentation structure		Presentation not given.	Very little awareness of presentation structure.	Some awareness of structure by using some signposting language and including an	Very good structure with a clear introduction and conclusion along with relevant	Impeccable presentation structure showing ability and competence beyond the	

				introduction.	signposting.	level.	
Challenging examiner		No indication of challenge or noticing of false assertion.	Candidate checks information but does not challenge.	Candidate challenges examiner but does not correct.	Candidate challenges examiner and corrects her/him.	Clear and immediate challenge and correction.	
Problem solving		No attempt made to tackle question.	Candidate shows very little acknowledgement of tasks to be completed.	Candidate addresses some tasks.	Candidate addresses almost all tasks.	Candidate attempts to address all tasks in a fluid manner.	
Asking questions		No questions asked at any point.	Relevant questions seldom asked.	Relevant questions sometimes asked.	Relevant questions asked frequently.	Frequent and relevant questions about a range of items in a clear and confident style.	
Empathy with audience		Candidate does not speak enough to give a grade.	Very little evidence of paraphrasing, elaboration of unknown concepts or explanation.	Some evidence of paraphrasing and elaboration but mostly verbatim repetition.	Very good awareness of audience by paraphrasing, explanation of unusual concepts and examples.	Outstanding ability to explain new or unusual ideas or concepts.	
Discourse management		Candidate does not speak enough to give a grade.	Very little acknowledgement of other candidate or ability to direct the conversation.	Candidate sometimes acknowledges other candidate and sometimes leads the dialogue.	Very well controlled conversation through relevant feedback to other candidate and frequently leading the conversation.	Excellent skill in directing, listening and focusing the conversation.	

Appendix 2

Questionnaire for EAP teachers

Introduction

This questionnaire is designed to gauge your attitudes as an EAP teacher towards a diagnostic test of speaking. It is hoped that, by gauging both teacher and student opinions, the test can be refined in such a way as to be a useful formative EAP assessment tool.

The test has been designed as part of my MSc in TESOL and Applied Linguistics dissertation project. The dissertation supervisor is Dr. XXXX

The test is in its pilot stages and therefore honest answers and any constructive criticism would be highly appreciated.

By completing this questionnaire voluntarily, you agree that any answers you give may appear in my dissertation and any other potential academic outputs (e.g. conference presentations and journal papers). However, your answers will remain completely anonymous. Your identity and institution will not be revealed under any circumstances.

If you have any concerns about this research, you have the right to complain to the dissertation supervisor or the Dean of the Faculty, University of YYYY, both of whose contact details are below.

If you would be willing to take part in a short follow up interview, please supply a contact email address at the end.

Please note: all questions are optional. If you cannot answer any question or do not wish to answer it, please move on to the next one.

Please get in touch with the researcher if you would like to see the findings from this study.

Many thanks for taking the time to complete this questionnaire!

Please click this link to access the test:

[\[Link\]](#)

Contacts:

(Researcher, MSc TESOL and Applied Linguistics student, University of YYYY)

[Email contact 1]

Dr. XXXX (Dissertation supervisor, University of YYYY)

[Email contact 2]

Professor ZZZZ (Dean of Faculty, University of YYYY)

[Email contact 3]

Part 1: Introduction

1. Please indicate your nationality:
2. Please indicate the country in which you currently work:
3. Please indicate your gender:
4. Did you use the test with your students? (*Yes/No*)
 - 5.1 If yes, please indicate if the students were: *undergraduate/postgraduate (non PhD)/postgraduate PhD/Other*
 - 5.2 If yes, how many students did you administer the test to?

Part 2: Overall test evaluation

5. Having either used the test or looked at it, please rate your level of agreement with the following statements: (*Strongly agree/agree/disagree/strongly disagree*)
 - 5.1 The test measures the types of speaking tasks that EAP students will need in their academic studies. (*Strongly agree/agree/disagree/strongly disagree*)
 - 5.2. The content of the test is relevant to EAP students. (*Strongly agree/agree/disagree/strongly disagree*)
 - 5.3 The test is easy to administer in the classroom. (*Strongly agree/agree/disagree/strongly disagree*)
 - 5.4 If the test was mandatory, it would require specific test preparation classes. (*Strongly agree/agree/disagree/strongly disagree*)
 - 5.5 Both candidates have the same opportunity to demonstrate their speaking ability. (*Strongly agree/agree/disagree/strongly disagree*)

5.6 The test allows students to demonstrate their academic speaking ability. (*Strongly agree/agree/disagree/strongly disagree*)

Part 3: Evaluation of test items

6. Question 1 focuses on more informal speech between students.

6.1 This is appropriate for an EAP test of speaking (*Strongly agree/agree/disagree/strongly disagree*)

7. Question 2 uses country profiles as content. Please indicate your general impression of countries as content in the test.

8. Please indicate your level of agreement with the following statements concerning question 2:

8.1 Countries are relevant content across students' subject disciplines. (*Strongly agree/agree/disagree/strongly disagree*)

8.2 Lesser known countries (i.e Guyana and Tanzania) facilitate discussion. (*Strongly agree/agree/disagree/strongly disagree*)

8.3 Uncommon words were deliberately included in the country profiles; this facilitates discussion. (*Strongly agree/agree/disagree/strongly disagree*)

9. Please indicate your general impression of the false assertion in the test.

10. Also in question 2 of the test you made a deliberate false assertion in your discussion with candidates. Please indicate your level of agreement with the following statements. (*Strongly agree/agree/disagree/strongly disagree*)

10.1 At least one candidate noticed the false assertion. (*Strongly agree/agree/disagree/strongly disagree*)

10.2 At least one candidate challenged me after giving the false assertion. (*Strongly agree/agree/disagree/strongly disagree*)

10.3 This false assertion is appropriate for an EAP speaking test. (*Strongly agree/agree/disagree/strongly disagree*)

11. Question 3 asks candidates to discuss a group assignment. Please indicate your level of agreement with the following statement:

11.1 This type of task is relevant to EAP students. (*Strongly agree/agree/disagree/strongly disagree*)

Part 4: Conclusion

12. How likely are you to use the test again? (*Very likely/likely/unlikely/very unlikely*)

13. Please indicate your general opinion of the test.

Appendix 3

Questionnaire for EAP students

This questionnaire is designed to gauge your attitudes as an EAP student towards a diagnostic test of speaking. It is hoped that, by gauging both teacher and student opinions, the test can be refined in such a way as to be a useful formative EAP assessment tool.

The test has been designed as part of my MSc in TESOL and Applied Linguistics dissertation project at the University of YYYY.

The test is in its pilot stages and therefore honest answers and any constructive criticism would be highly appreciated.

By completing this questionnaire voluntarily, you agree that any answers you give may appear in my dissertation and any other potential academic outputs (e.g. conference presentations and journal papers). However, your answers will remain completely anonymous. Your identity and institution will not be revealed under any circumstances.

If you have any concerns about this research, you have the right to complain to the dissertation supervisor or the Dean of the Faculty, University of YYYY, both of whose contact details are below.

If you would be willing to participate in a very short follow up interview, please supply a contact email address at the end of the questionnaire.

Please note: all questions are optional. If you cannot answer a question or do not wish to answer it, please move to the next one.

Please contact the researcher if you would like to see the findings from this study.

Many thanks for taking the time to complete this questionnaire!

Contacts:

(Researcher, MSc TESOL and Applied Linguistics student)

[Email contact 1]

Dr. XXXX (Dissertation supervisor)

[Email contact 2]

Professor ZZZZ (Dean of Faculty)

[Email contact 3]

Part 1: Introduction

1. Please indicate if you are: pre-sessional undergraduate/pre-sessional postgraduate (non PhD)/pre-sessional postgraduate PhD/Other
2. Please indicate your nationality
3. Please indicate the country in which you currently study.
4. Please indicate if you are: male/female

Part 2: Overall test evaluation

5. Please indicate your agreement with the following statements:

5.1 The test measures the types of speaking tasks that are relevant to my academic studies. (*Strongly agree/agree/disagree/strongly disagree*)

5.2 The content of the test is relevant to my academic studies. (*Strongly agree/agree/disagree/strongly disagree*)

5.3 The test does not require any special resources.

*(Strongly
agree/agree/disag
ree/strongly
disagree)*

5.4 I had the same
chance as the other
candidate to
demonstrate my
speaking ability.

*(Strongly
agree/agree/disag
ree/strongly
disagree)*

5.5 If I did the test
again, I would like
to attend specific
test preparation
classes. *(Strongly
agree/agree/disag
ree/strongly
disagree)*

5.6 The test allowed me to demonstrate my academic speaking ability. *(Strongly
agree/agree/disagree/strongly disagree)*

Part 3: Evaluation of test items

6. In question 1 of the test you spoke to the other candidate in a more informal way.

6.1 This is relevant to my academic speaking ability. *(Strongly agree/agree/disagree/strongly
disagree)*

7. In question 2 of the test the examiner deliberately said something false in the discussion. This is called a 'false assertion'.

7.1 Did you notice this? *(Yes/no)*.

7.2 If you answered yes, did you challenge the examiner? *(Yes/no)*.

7.3 To what extent do you agree with the following statement: The false assertion is helpful in developing my academic speaking abilities. . *(Strongly agree/agree/disagree/strongly disagree)*

Part 4: Conclusion

8. To what extent do you agree with the following statement:

8.1 I would like to do the test again. *(Strongly agree/agree/disagree/strongly disagree)*

9. Please give your general opinion of the test, including any criticism.

Appendix 4

Changes to questionnaires

EAP Teachers

Please note: since question numbers changed from the first to the second draft, only revised question numbers have been included.

Original question	Revised question	Rationale
If yes, who were the students? Pre-session undergraduate / pre-session postgraduate / other (please specify)	4a. If yes, please indicate if the students were: <i>undergraduate/postgraduate (non PhD)/postgraduate PhD/Other</i>	To account for being postgraduate PhD and postgraduate masters (or similar).
The test measures the types of speaking tasks that EAP students will need in their courses. Strongly agree/agree/disagree/strongly disagree	5.1 The test measures the types of speaking tasks that EAP students will need in their academic studies. Strongly agree/agree/disagree/strongly disagree	To standardise the wording
The test would influence my teaching (i.e ‘wash back’ effects). Strongly agree/agree/disagree/strongly disagree	5.4 If the test was mandatory, it would require specific test preparation classes.	To avoid ambiguity that may arise if teachers did not know the ‘washback’ concept.
The test could be replicated. Strongly agree/agree/disagree/strongly disagree	5.5 Both candidates have the same opportunity to demonstrate their speaking ability. (<i>Strongly agree/agree/disagree/strongly disagree</i>)	To ask a more specific question concerning the test’s reliability.
Do you think this use of countries is relevant across EAP students’ subject disciplines? YES/NO	8.1 Countries are relevant content across students’ subject disciplines. (<i>Strongly agree/agree/disagree/strongly disagree</i>)	To allow a more quantifiable response, Likert scales were used.
Lesser known countries were deliberately chosen. Do you think	8. 2 Lesser known countries (i.e Guyana and Tanzania) facilitate	To allow a more quantifiable

<p>this facilitates discussion? YES/NO</p>	<p>discussion. (<i>Strongly agree/agree/disagree/strongly disagree</i>)</p>	<p>response, Likert scales were used.</p>
<p>Some lesser known words were included in the country profiles. Do you think this facilitates discussion? YES/NO</p>	<p>8.3 Uncommon words were deliberately included in the country profiles; this facilitates discussion. (<i>Strongly agree/agree/disagree/strongly disagree</i>)</p>	<p>To allow a more quantifiable response, Likert scales were used.</p>
<p>Did the candidate(s) notice this? YES/NO</p>	<p>10.1 At least one candidate noticed the false assertion. (<i>Strongly agree/agree/disagree/strongly disagree</i>)</p>	<p>To avoid ambiguity of whether one or two candidates was being referred to. A Likert scale was also used to gain more quantifiable data.</p>
<p>Did the candidate(s) challenge you? YES/NO</p>	<p>10.2 At least one candidate challenged me after giving the false assertion. (<i>Strongly agree/agree/disagree/strongly disagree</i>)</p>	<p>To avoid ambiguity as to whether one or two candidates are being referred to.</p>
<p>Do you think this feature is appropriate? YES/NO</p>	<p>10.3 This false assertion is appropriate for an EAP speaking test. (<i>Strongly agree/agree/disagree/strongly disagree</i>)</p>	<p>To gain more quantifiable data through a Likert scale.</p>

Appendix 5

Changes to questionnaires

EAP students

Please note: since question numbers changed from the first to the second draft, only revised question numbers have been included.

Original question	Revised question	Rationale
Please indicate if you are: pre-sessional undergraduate / pre-sessional postgraduate / other (please specify).	1. Please indicate if you are: pre-sessional undergraduate/pre-sessional postgraduate (non PhD)/pre-sessional postgraduate PhD/Other	To distinguish between postgraduate PhD and postgraduate masters' students.
The test allowed me to demonstrate my speaking ability. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	5.6 The test allowed me to demonstrate my academic speaking ability. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	To be more specific by adding the word 'academic'.
I understood the test questions. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	5.3 The test does not require any special resources. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	To more accurately measure student attitudes towards test practicality.
The feedback from the test was helpful. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	5.5 If I did the test again, I would like to attend specific test preparation classes. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	To focus on a different aspect of test impact.
The types of tasks are similar to what will be expected during my university course. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	5.1 The test measures the types of speaking tasks that are relevant to my academic studies. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	To be more specific in the question wording.
Not included	6.1 In question 1 of the test you spoke to the other candidate in a more informal way. This is relevant to my academic	To gain insight on beliefs about informal language use in EAP.

	speaking ability. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	
Do you think this feature is useful for your speaking abilities? <i>YES/NO</i>	7.3 To what extent do you agree with the following statement: The false assertion is helpful in developing my academic speaking abilities. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	To gain greater insight and provide respondents with a greater variety of options.
Would you like to do the test again? <i>YES/NO</i>	3. To what extent do you agree with the following statement: I would like to do the test again. <i>(Strongly agree/agree/disagree/strongly disagree)</i>	To gain greater insight and provide respondents with a greater variety of options.

Appendix 6

Email distributed to all subscribers on the BALEAP mailing list

Dear All,

I have developed a diagnostic speaking test for EAP students as part of my MSc TESOL dissertation at the University of YYYY. The test is designed for teachers to administer in the classroom.

The test is original and contains several novel features. As part of my research, I am hoping to obtain EAP teacher feedback on the test and its design. Teachers can use the test and then complete a short questionnaire giving some feedback or merely have a look at the test and complete the questionnaire. The questionnaire takes less than 10 minutes to complete.

Since testing in EAP, and in particular testing speaking, seems to be a burning issue I hope my test design will add something to the ongoing debate. Teachers are free to use the test as they wish and it will also hopefully provide some useful feedback to their students.

The link to the questionnaire is below; the link to the test itself (which also contains instructions and mark sheets) is contained in the questionnaire introduction.

[Website address](#)

Please feel free to contact me regarding any aspect of my research or the test itself.

Many thanks,

Student, MSc TESOL and Applied Linguistics

[Email contact](#)