

English Language Assessment in the Colleges of Applied Sciences in Oman: Thematic Document Analysis

Fatma Al Hajri¹

¹Colleges of Applied Sciences-Sur, Ministry of Higher Education, Oman

Correspondence: Fatma Al Hajri, Ministry of Higher Education, P.O. Box 82, Airport heights, PC 112, Oman.
E-mail: fatimaalhajri@gmail.com

Received: December 10, 2013 Accepted: January 2, 2014 Online Published: February 12, 2014

doi:10.5539/elt.v7n3p19 URL: <http://dx.doi.org/10.5539/elt.v7n3p19>

Abstract

Proficiency in English language and how it is measured have become central issues in higher education research as the English language is increasingly used as a medium of instruction and a criterion for admission to education. This study evaluated the English language assessment in the foundation Programme at the Colleges of Applied sciences in Oman. It used thematic analysis in studying 118 documents on language assessment. Three main findings were reported: compatibility between what was taught and what was assessed, inconsistency in implementing assessment criteria, and replication of the General Foundation Programme standards. The implications of the findings on national and international higher education are discussed and recommendations are made.

Keywords: higher education, national policies, language education, document analysis

1. Introduction

In the Colleges of Applied Sciences (CAS), the English language was chosen to be the language of instruction when various English speaking higher education “policy entrepreneurs”, as Ball (1998) calls them, were invited to put forward their proposals and plans for the six amalgamated Colleges. In 2006, the Ministry of Higher Education, under which the Colleges operated, signed a contract with Polytechnics International New Zealand (PINZ) to conduct a needs analysis of the labour market and recommend the future academic programmes of the colleges. The programmes offered by the colleges currently, as a result of the PINZ report, are Information Technology, Design, International Business Administration and CS. This approach to creating new HEIs has been criticised for being totally foreign to the local cultures; Donn and Al-Manthri (2010, p. 24) argue that “they [the Gulf countries] have little control, other than as purchaser and consumer, over the language or the artefacts of the language”.

When the programmes the colleges would offer were agreed upon, New Zealand Tertiary Education Consortium was contracted to provide the curriculum as well as part of the assessment and other services. The first batch of the students had to go through an English language preparation programme (i.e., foundation programme) for almost an academic year before qualifying to take the academic courses in English. The assessment documents used in the English Language programme display the foreignism of the programme created by the tensions between the national needs and international requirements of the language programme.

2. The Foundation Programme

In Oman, almost 80% of high school graduates admitted to higher education take English language courses in the Foundation Programme (FP) before embarking on academic study (Al-Lamki, 1998). “The FP is a pre-sessional programme that can be considered an integral part of almost all of the HEIs in Oman. Its general aim is to provide students with the English language proficiency, study skills, computer and numeracy skills required for university academic study (OAAA, 2009)” (Al Hajri, in press).

Table 1. English language courses in the foundation programme and their approximate equivalent levels in IELTS

Equivalent in IELTS	Foundation Programme levels	Courses	Weekly contact hours (Hrs.)	Total Hrs.
IELTS 3.0 or below	Level C	GES	11	20
		AES	9	
IELTS 3.5	Level B	GES	11	20
		AES	9	
IELTS 4.0	Level A	GES	11	20
		AES	9	
IELTS 4.5	Entry to First Year	EAP	10	10

*Modified from *Colleges of Applied Sciences Prospectus* (2010, p. 33).

As shown in Table 1, FP consists of two main courses, the General English language (GES) and Academic English Skills (AES) which are allocated twenty hours per week. In addition FP includes two hours of mathematics and/or computer skills courses in each semester. In this paper, FP refers to the English language courses.

3. Language Assessment in the Foundation Programme

The academic regulations of CAS state that 50% of a course scores should be allocated to CA and the other 50% to the final test (CAS, 2010e).

Table 2. Assessment instruments in the foundation programme courses (Al Hajri, in press)

Course	Assessment instruments	% Course total	% Foundation Programme total
General English Skills	Mid-term Test	40%	50%
	Final Test	60%	
Academic English Skills	Presentation	50%	50%
	Report	50%	

In the FP, students take two courses in which they undergo two different assessment instruments. Table 2 shows that assessment in the GES course includes a mid-term test and a final test, whereas assessment in the AES course includes writing a report and presenting it orally. Students are required to obtain 50% of the total marks in each course.

4. Study Questions

This study presents and discusses the results obtained from document analysis conducted as part of a more comprehensive mixed method study on FP assessment and this paper aims at responding to the four following questions:

- 1) What processes and procedures were followed in writing and implementing the assessment instruments, as depicted by the official documents?
- 2) What were the differences between the 'continuous assessment' model used in the Academic English Skills course and the 'test' model used in the General English Skills course in terms of effectiveness, accuracy, and preferences of teachers and students?
- 3) What types (criterion/norm-referencing) of assessment were used? And how?
- 4) What were the national policies on teaching and assessing language that influenced assessment in Oman? And how does FP assessment correspond to these policies?

5. Background on the Role of Documents in the Foundation Programme

The documents analysed in this study vary in type, length, accessibility and implementation. Most of them were

centrally issued by the Directorate General of the Colleges of Applied Sciences (CAS), some were issued by the Oman Academic Accreditation Authority (OAAA), and others by the Ministry of Higher Education.

Table 3. A Selection of documents relating to teaching and assessment of the FP English language course

Type	Documents
General	Oman Academic Standards for General Foundation Programmes Colleges of Applied Sciences: Academic Regulations Student Guide for Colleges of Applied Sciences (2011/12) Academic Audit Reports on Colleges of Applied Sciences in Sohar, Ibri and Salalah
Teaching	Foundation Programme: 2010-11 Course Specifications for Foundation English Headway Academic Skills (Level 2) Headway Plus (Intermediate) Essay and Presentation Guidelines Foundation Year Academic Calendar
Assessment	CAS English Department Assessment Handbook Foundation Year – Level A Academic Skills Project & Presentation Topics Mid-term and Final Tests for Level A Foundation English Assessment Policies: English Department October 2011 English Department Anti-Plagiarism Procedures: Student plagiarism V3, 02/11 Marking Scales for Tests and Projects

The types of documents can be categorised in terms of their focus into general documents, teaching documents and assessment documents. About 118 documents were investigated in this study, varying in length from one page to 50 pages. Table 3 displays a sample of these documents.

The accessibility of these documents to FP teachers depends on their position and their target audience. Some of the general documents were accessible to the heads of departments, but not the teachers; others were accessible to all and could be retrieved from the Internet. The general documents could be claimed to be unnecessary for the teachers as they mostly included policies, regulations or audition reports, and consequently, they were not distributed to teachers, though they were available online. The teaching documents were intended to be supplied to every teacher on the FP. It was the responsibility of the course coordinators in each college to supply the teachers with these documents, which were exclusively accessed online by the coordinators. This means that the number of teaching documents the teachers received was bound to how much and how widely a coordinator disseminated these materials. Similarly, circulation of the assessment documents depended on the assessment coordinators at the colleges who had exclusive online access to these materials. All of the documents on assessment tasks, specifications and marking scales were supposed to be shared with the teachers. Current and previous tests however, were accessed by the assessment coordinators only, to allow a possible recycling of the test tasks.

The level of teacher participation in and implementation of the FP English course documents also differed according to the document types. In general, not all teachers participated in writing the documents, including the tests and assessment tasks. Only the assessment coordinators, who taught a lower number of hours, participated in writing the tests. In regard to the implementation of policy documents and marking scales, there was no accountability system in place. However, there were standardisation workshops held for marking the writing task of the General English Skill (GES) final test, and a two-rater policy was followed in evaluating the students' speaking skills in the GES interview; no similar workshops were conducted on the standardisation of marking the Academic English Skills (AES) assessment.

In carrying out the document analysis, I was trying to understand in a factual way the plans and intentions and

was deliberately using a problem centred approach to find possible contradictions.

6. Document Analysis

The approach to document analysis was thematic analysis that is ‘a form of pattern recognition’ (Bowen, 2009, p. 32). Although in the design of this study a critical hermeneutics approach was intended to guide the document analysis, it was found to be impractical for the purposes of the study and types of documents collected. Critical hermeneutics as developed by Philips and Brown (1993) and Forster (1994) focused on both the context of the documents within which they were produced and the point of view of the author in generating common themes. Linking the themes to the context and authors’ views was not chosen in this study for two reasons. First, the document analysis was one of four sources of data in a more comprehensive study conducted for the requirements of a Doctorate Degree; therefore, it was felt that applying similar codes to those generated by the interviews and focus groups would facilitate integrating data (Bowen, 2009). Second, the author’s views and context of the documents could not be identified for all the collected documents (e.g., student marks, and task specifications). Therefore thematic analysis was employed in document analysis to facilitate comparing and contrasting the results from different data sources. This comparison is intended to reveal the reality of what is presented in the documents. Atkinson and Coffey (2004) argued that documents are written with hidden purposes in mind and they could suppress some realities if they were to be displayed in public, so the writers warned that

We cannot ... learn through written records alone how an organization actually operates day by day. Equally, we cannot treat records - however “official”- as firm evidence of what they report (Atkinson & Coffey, 2004, p. 58).

To ease retrieving coded extracts from this large number of documents, Atlas ti. (i.e., a qualitative data analysis tool, see Figure 1) was used. The documents were uploaded into the software which was strictly used only to organise the documents and codes for faster retrieval.

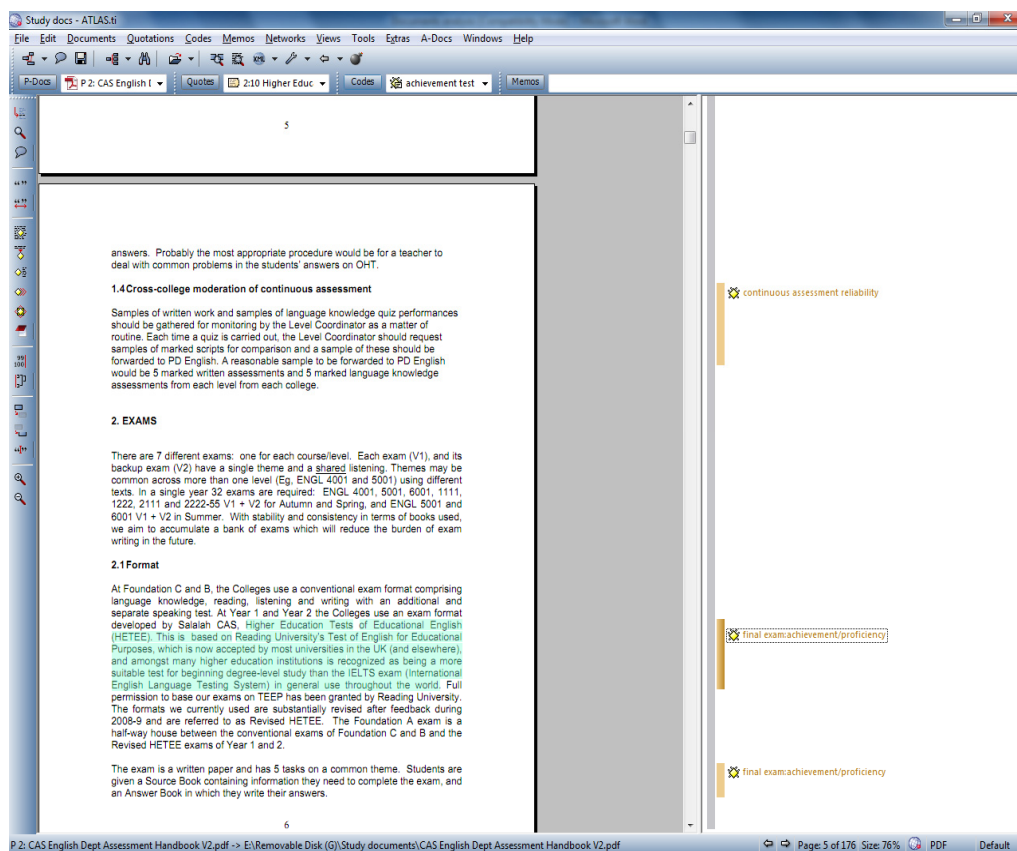


Figure 1. Assigning codes to texts using Atlas ti

The analysis process went through several steps to generate themes that embodied the main issues on the quality of assessment writing and implementation in the FP. These steps are described below:

- 1) Initial reading and highlighting of possible important points.
- 2) Secondary reading that included forming a list of codes that either emerged while reading or were used in the interviews and focus group analyses.
- 3) Refining the codes by excluding the less common ones and the ones that were irrelevant to the subject of the study.
- 4) Uploading the codes to Atlas ti. The figure below shows a document in the coding process. The codes are on the right hand side and the document is on the left hand side. When a code is selected the linked extracts become highlighted.
- 5) Reading the documents again prior to assigning the selected codes.
- 6) Coding the documents. Returning to the questions of the study to focus the codes.
- 7) Reading the extracts and organizing them into themes. Going back to the original texts to check if themes are appropriate and comparing them to the themes generated by the other methods to ensure that similar themes were focused upon in the analysis.
- 8) Writing up the results based on the themes found.

7. Results

The results are categorised into four main themes: (1) conflicts and tensions between criterion-referenced and norm-referenced assessment, (2) compatibility between what was taught and what was assessed, (3) inconsistency in implementing assessment criteria, (4) replication of the academic standards in the FP course specifications. The first, second and third themes focused on the design, implementation and marking of the assessment tasks respectively (i.e., a micro perspective). The fourth theme focused on the evaluation of FP assessment in the context of the national standards of the FP in Oman and its suitability for the language requirements of the FY academic courses (i.e., macro prospective). These themes emerged after implementing the coding process explained in section 4.7.1.

7.1 Conflicts and Tensions between Criterion-Referenced and Norm-Referenced Assessment

Generally assessment instruments are used for either norm-referenced, or criterion-referenced purposes depending on stake-holders' or institutions' needs. Norm-referenced testing (NRT) "relates one candidate's performance to that of the other candidates. We are not told directly what the student is capable of doing in the language" (Hughes, 2003, p. 20). Criterion-referenced tests (CRT) aim to "classify people according [to] whether or not they are able to perform some task or set of tasks satisfactorily" (Hughes, 2003, p. 21).

The English language components of the FP consisted of two courses: AES and GES. At the time of this study, GES assessment included a midterm test and a final test that were centrally written, whereas AES assessment included report writing and an oral presentation of the report. Investigation of the official documents on constructing the GES tests appeared to show that there was a sort of incongruity among different official documents about whether the purpose of these tests was norm-referencing or criterion-referencing. For example, the test writing instructions in the *English Department Assessment Handbook* (2010) advised using what could be considered norm-referenced techniques in writing test items and analysing student scores. However, the *CAS Regulations, General Foundation Programme Standards (GFPs)* and *English Department Course Specifications* all stated that the tests should aim at assessing students' abilities to achieve set outcomes and, should be using criterion-referenced achievement tests. The policy documents of the Colleges and of the national accreditation institution namely *CAS Academic Regulations* and *Oman Academic Standards for General Foundation Programs*, clearly mandated that assessment instruments should have the traits of a criterion-referenced assessment not a norm-referenced one. This is explicitly stated in the extracts below.

Normally a final grade in any given course is based on continuous evaluation of the achieved Learning Outcomes. This implies therefore that assessment is determined more by the fulfillment of stated criteria rather than by solely comparative achievement within a class (CAS, 2010a, p. 15).

All assessment shall be criteria based (i.e., based on the learning outcome standards) and not normative references. Arbitrary scaling of results (for example, ensuring a certain percentage of students passes by moving the pass/fail point down the scale of student results) shall not be permitted (OAAA, 2009, p. 8).

However, the English department's documents seemed to give conflicting guidance. Although, these documents stated that the tests aimed at evaluating students' mastery of a set of learning outcomes, and thus implied that they should be criterion-referenced, the test writing and analysing instructions entailed using norm-referenced

methods that compared the students' performances to each other, as in this extract:

Item analysis will be carried out by the Assessment Team based on samples of marks from a single college. This analysis involves counting the numbers of correct answers given for each item by the sample population. From this analysis a number of conclusions can be drawn:

- 1) Items which nobody gets right or items which everybody gets right are to be marked for deletion or alteration in subsequent versions of the test.
- 2) Items where 25% or less of the population gets the correct answer need to be investigated: if the 25% of the sample getting the answer right are also the 25% highest scoring students, this is a positive indicator. If no such correlation is found, the item needs to be marked for deletion or alteration in subsequent versions of the test ... Such items should be recorded to build up a bank of bad test items in order to guide future test writing (CAS, 2009, p. 20).

This was also apparent in the following instructions in the newer version of the same document:

Preliminary analysis of marks: This should include (a) a check on relative scores for representative students i.e., students who are recognised to be high-achieving, middle-range, low-achieving. If these students are placed in 12 more or less the order teachers would expect, this is a positive indicator (b) a check on relative scores for groups. Again this relates to recognised prior achievement: if groups perceived to be achieving at the same levels score roughly the same, this is a positive indicator (CAS, 2010c, p. 12).

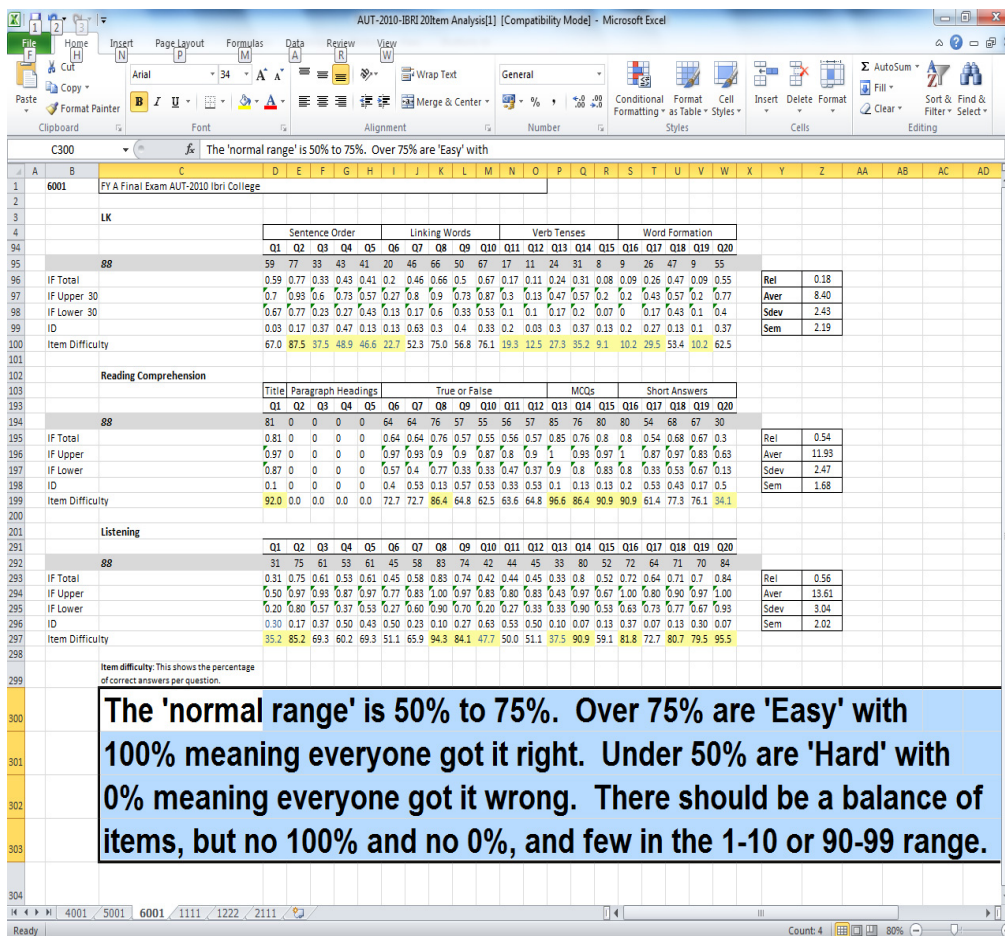


Figure 2. Guidance for the FP teachers on tests item analysis in 2010

Also, Figure 2 shows that the process of item analysis focuses on selecting the test items using the normal distribution curve, to ensure that most of the population fall in the middle range of the distribution.

Though the GES tests did not comply with CAS or OAAA policies on implementing criterion-referenced tests, they did follow the policies on testing achievement, not proficiency. It is stated in the *English Department*

Assessment Handbook (2009, p. 3) that “the purpose of the test is to show achievement”. Hughes (2003, p. 13) says that achievement tests “establish how successful individual students ... have been in achieving objectives” and identifies the aim of proficiency tests to be “measure[ing] people’s ability in a language regardless of any training they may have had in that language” (Hughes, 2003, p. 11). It seems that CAS students were generally assessed on a predetermined set of outcomes rather than on general proficiency in certain skills or abilities, as the policy makers intended.

On the other hand, the AES assessment instruments seemed to be designed to evaluate the students’ language abilities using criterion-referenced and achievement measures as recommended in CAS regulations and OAAA standards. This was deduced from reviewing the specifications of the AES report and presentation that assessed FP students based on their achievement of a certain set of criteria, and was also expressed in the following extract.

Continuous assessments are designed to provide teachers and students with an on-going measure of achievement so that they can both adjust expectations and level of input (CAS, 2010c, p. 4).

7.2 Compatibility between What Was Taught and What Was Assessed

By comparing and contrasting the focus of assessment instruments with the focus of the taught materials, this section sheds some light on what was claimed to be assessed and what was actually assessed in each course by comparing textbooks, course specifications, test specifications and papers, and continuous assessment specifications and tasks. This part of the study followed an objective based model of evaluation which investigates if the objectives of a programme have been met.

Table 4. Textbooks and assessment in AES and GES courses^a

FP course	Textbooks	Assessment components		% Course total	% English FP total
GES	<i>New Headway Plus Intermediate</i> <i>New Headway Plus Intermediate Workbook</i>	Midterm test	Language knowledge	10%	50%
			Reading	20%	
			Listening	20%	
			Speaking	20%	
			Writing	30%	
		Total	40%		
		Final test	Language knowledge	10%	
			Reading	20%	
			Listening	20%	
			Speaking interview	20%	
Writing	30%				
Total	60%				
AES	<i>New Headway Academic Skills (Level 2)</i>	Presentation		50%	50%
		Report		50%	

^a Taken from (CAS, 2010b, p. 19).

Table 4 displays the textbooks and assessment tasks used in each course. It can be seen from the table that GES assessment consisted of tests, while AES assessment consisted of performance assessment tasks (i.e., a report and presentation).

7.2.1 Compatibility in GES Learning Outcomes, Taught Materials and Test Tasks

Analyses of GES and AES documents are presented separately. First the GES course materials, textbooks, tests, and scales were examined to understand what the students were supposed to be taught and what was supposed to be included in the tests according to official documents. An initial comparison of the intended GES learning outcomes, as stated in the *Course Specification for Foundation English*, and the GES test specifications, as stated

in the *English Department Assessment Handbook*, revealed a very close resemblance, suggesting that most of the skills the students should master by the end of the course seemed to be measured by the tests, if the students' met the specifications. For example, the *Course Specification for Foundation English* stated that "by the end of the course, students should be able to read texts of up to 600 words, with a Flesch test readability score of 85%, with gist, main points and detailed comprehension" (CAS, 2010c, p. 16). This objective was found to be addressed in the *English Department Assessment Handbook*, which stated that the reading passage used in the final test should be "500-550 words of length and of around 80% of readability" (CAS, 2010c, p. 20). From this example and several others, it can be inferred that the GES test specifications seemed to correspond to the learning outcomes by using tasks of appropriate levels. It can also be suggested that since GES test tasks focused on covering most of the learning outcomes, GES tests fulfilled the requirements of content validity (i.e., the extent to which a test represents all facets of a content domain).

Despite the general compatibility between the course learning outcomes and the test specifications, an analysis of the GES course textbook (i.e., *New Headway Plus Intermediate*) showed that its content, especially its tasks, were of a shorter length than those suggested by the course learning outcomes and test specifications. For example, the reading scripts provided in the textbook seemed to be significantly shorter than the 600 word passages used in the test. Also, the course specifications stated that students should be able to produce 350 word written scripts, yet the writing tasks in the textbook were based on shorter passages. This suggests that the students possibly lacked sufficient and appropriate input to meet the test tasks' requirements. The taught materials were of a shorter length than of that stated in the course learning outcomes and test specifications.

That being said, most of the general topics mentioned in the GES textbook (e.g., talking about films, and cities) were systematically similar to the topics the learning outcomes and test specifications addressed. This was true for each of the reading, writing, and speaking skills, but not for the listening skill.

Although the assessed learning outcomes of the listening skill matched those of the textbook, the test specifications introduced an unfamiliar listening genre to the students (i.e., listening to lectures). The test specifications stated that two listening tasks should be used: (1) a dialogue between two people, and (2) a lecture. However, the lecture genre did not occur in either the textbooks or the listening skill learning outcomes of FP course specifications. Listening to a lecture could be more difficult for the students as a genre; it is a monologue which usually lacks social interaction cues. Though some might argue that this type of listening task is more authentic, it is different to what the students were taught in class (e.g., discussion, role-play and description) and perhaps more complex. After the midterm test was administered in Spring 2011, the issue of the listening task difficulty came up in several focus groups. Likewise, the difficulty of the listening component of the test was not expressed only by the students, it was also acknowledged in the *English Department Assessment Handbook*, "listening is the most difficult task for students" (2010c, p. 8). This reoccurrence of instances where the listening tasks were deemed to be difficult for the students implies a consensus on the inappropriateness of the listening task level or type.

7.2.2 Learning Outcomes, Taught Materials and Assessment Tasks in the AES Course

As in the case of the GES tests, the specifications for the report and presentation task used in the AES assessment closely mirrored the intended AES learning outcomes, but again the assigned textbook seemed unable to fulfil the ambitious stated specifications of the assessment and learning outcomes. The learning outcomes in the *Course Specifications for foundation English* included statements such as, "produce a written report of a minimum of 500 words" (CAS, 2010b, p. 19), and "read an extensive text of around 1,000 words broadly relevant to an area of study and respond to questions that require analytical skills, e.g., prediction, deduction, inference" (CAS, 2010, p. 19). However, the course textbook, *New Headway Academic Skills (Level 2)*, included reading passages of a maximum length of 600 words and assigned writing activities of 250 word essays. A comparison of the language difficulty levels of the textbook materials and those of the learning outcomes and test specifications reveals considerable differences between them indicating that test specifications might generate test tasks of a more difficult level than those experienced by students in the classroom.

Instructions for report writing and presenting in AES course: (English Department, 2011, p. 1)

- 1) Students are required to complete a project which involves some library, Internet and real-world research (e.g., interviewing people), a presentation and a report.
- 2) Students should choose a topic from the list below [the list was attached to the instruction sheet]. The topics are based on the subjects the students will study this semester.
- 3) The subjects are quite wide so the student and teacher should agree the actual scope/title of the report.

4) Students should not write about Oman or Omani related topics. As part of their project they are required to do research about a *new* topic.

The report should be around 500 words and the presentation should be at least 5 minutes. Each part represents 50% of the marks.

In order to understand the nature of what seems to be assessed using performance based tasks (e.g., a report and a presentation), studying the tasks alone was not enough. The marking scales had to be considered too as they determined the focal points of an assessment through the criteria used. In this study, the band descriptors of the AES learning outcomes and of the marking scales were compared and a discrepancy was found between what was intended to be taught and what seemed to be assessed. Interestingly, this discrepancy was found only between the *writing* learning outcomes and *writing* marking scale descriptors but not between *speaking* learning outcomes and the *speaking* scale descriptors. Before a fuller description, it seems necessary to first clarify the nature, structure and specifications of the AES assessment tasks: report and presentation. Box 1 displays the instructions which teachers were supposed to share with their students on the AES assessment.

Table 5. Comparison of AES writing learning outcomes and marking scale descriptors

Comparisons	Writing learning outcomes	The highest level of the marking scale
Corresponding areas	Produce a written report of a minimum of 500 words showing evidence of research, note taking, review and revision of work, paraphrasing, summarising, use of quotations and use of references. Cite sources according to the APA system.	<i>All outlines and drafts completed and submitted on time.</i> <i>Student has actively tried to implement all changes suggested by teacher.</i> <i>Majority of the essay is in the students own words and credit is given when others' work is used.</i> <i>Meets minimum word limits.</i>
	Plan and execute a piece of writing by moving through a series of process stages. Use mind-maps to brainstorm content for writing. Use linking words to show logical organisation within and across sentences.	Addresses chosen topic directly; coverage is fairly comprehensive; little irrelevance. Essay structure used includes introduction, conclusion etc.
Conflicting areas	Proof-read effectively focusing on a range of surface features. Complete applications forms. Reformulate phrases from a sentence. Paraphrase sentences from a text. Summarise paragraphs from a text. Use pronouns to avoid repetition. Use modal verbs (e.g., may, could) and adverbs of possibility (e.g., possibly). Transfer information from graph to text and text to graph.	No corresponding descriptors

As has been noted earlier, the speaking learning outcomes in the AES course closely resembled the presentation marking scale. Table 6 displays the similarities between the speaking learning outcomes and the highest level of the speaking scale descriptors by placing corresponding learning outcomes and descriptors next to each other.

Table 6. Comparison of AES speaking learning outcomes and scale descriptors

Comparisons	AES speaking learning outcomes	The highest level of the presentation marking scale
Corresponding areas	Prepare and deliver a talk of at least five minutes. Use library resources in preparing the talk, speak clearly and confidently, make eye contact and use body language to support the delivery of ideas. Respond confidently to questions.	Gets the attention of the audience: highlights objectives of presentation
	Address questions from the audience.	Postures, gestures and movement enhance presentation.
	Plan and conduct a presentation based on information from written material, interviews, surveys, etc.	Complete understanding of topic. Clear evidence of independent study. Able to effectively answer any questions on the topic.
	Tailor content and language to the level of the audience.	
	Maintain some eye contact with audience.	
	Outline and define main concepts.	Presentation well organised with a logical flow of information
	Follow a presentation format.	
Conflicting areas	Use presentation language (discourse markers etc.).	
	Achieve the key aim of informing the audience.	Topic was covered thoroughly and concisely. No important information missed
	Observe time restrictions in presentations.	Reiterates key points: pulls the entire presentation together effectively.
	Organise and present information in a logical order at a comprehensible speed.	Uses allotted time fully.
	Speak in a clearly audible and well-paced voice.	Few pronunciation errors: delivery is clear.
Conflicting areas	Make use of audio/visual aids when giving oral presentations.	Few grammatical errors; none of which cause confusion.
	Invite constructive feedback and self-evaluate the presentation.	A wide range of appropriate vocabulary, correctly used.

The focus of the scale used to mark the written report was found to be different from that of the writing learning outcomes of the AES course; these differences were apparent when the learning outcomes of the AES writing skills were placed next to the highest level of the writing marking scale as shown in Table 6. It can be seen from the table that four of the six criteria in the scale evaluated the structures and procedures of writing an essay (i.e., word count, plagiarism and implementing suggested changes). All of these four italicised criteria correspond in focus with two learning outcomes of the writing skill in the left hand side of the table. In the scale, there were only two criteria that focused on the content of the report, namely the fifth and sixth points: “addresses chosen topic directly” and “essay structure used includes introduction, conclusion ...etc.” Areas such as linguistic knowledge (e.g., using pronouns or modal verbs), and stylistic knowledge (e.g., using paraphrases) were listed in the learning outcomes but were overlooked by the marking scale. It can be inferred from the marking scale that regardless of the quality of a written piece, a student could easily score a high score if he submitted on time, his report was within the word limit, he wrote it by himself, and he followed a teacher’s suggestions.

In general, the comparison of AES assessment documents revealed instances of what could be regarded as an imbalance amongst the learning outcomes, textbook materials and marking scales in all of the four skills. The learning outcomes of the writing skill were of a higher difficulty level than the textbook writing activities, and the focus of the writing marking scale differed from that of the learning outcomes. Similarly, the reading outcomes were of higher difficulty level than the reading activities in the textbook; however, there was not any assessment task on this skill in the AES course. The speaking learning outcomes were not covered by the textbook, but they were almost comprehensively represented in the marking scale, unlike the listening ones which were not covered by the textbook and were not assessed.

The attempt to understand how tests and assessment tasks functioned in the GES and AES courses by exploring the larger picture that encompassed the courses' learning outcomes, textbooks, assessment instruments and marking scales showed that what was stated to be assessed did not always correspond with what was actually assessed.

7.3 Inconsistency in Implementing Assessment Criteria

The reliability and consistency of assessment instruments in measuring intended English language skills are crucial to effectiveness and validity of language programme assessment. Therefore educational institutions usually record how reliable their assessment instruments are and how consistency in using certain measures should be realised. Accreditation and quality assurance agencies usually urge academic institutions to (1) use reliable measures of achievement, and (2) state the process used to insure consistency in applying these measures. The General Foundation Programmes standards (GFP) as set by the OAAA emphasise the necessity of putting in place appropriate procedures to ensure the required level of moderation and standardisation in language assessment. The extract below addresses Higher Education Institutions (HEIs):

HEIs must have appropriate internal quality controls for its assessment processes. These must include, at least, internal moderation by faculty of examination papers and of marked work prior to the issuance of results, and a transparent appeals process for students (OAAA, 2009, p. 8).

In line with the OAAA standards for moderation and standardisation, CAS regulations included an article on forming a committee responsible for ensuring that standardisation policies within and across the six Colleges are met.

The aim of the [Examiners] Committee is to:

- 1) Ensure consistent standards of quality within the program and across all Colleges, by reviewing the performance for each student enrolled into the program;
- 2) Ensure that all evaluation and grading is performed in a fair and equitable manner, and in accordance with these Regulations (CAS, 2010a, p. 15).

The English language Department at CAS, following the guidelines of OAAA and CAS on standardisation and moderation of assessment, issued three policy documents in 2009, 2010 and 2011 respectively. Each of the documents implied that the previous one had fallen short of fulfilling standardisation requirements; it was stated that "unfortunately, this [standardisation] approach has presented severe reliability problems because of varied levels of challenge and it has also meant an excessive workload for coordinators" (CAS, 2010c, p. 5). The changes in the standardisation and moderation policies have been tracked from 2009 to 2012; these changes are listed in Appendix 1 to reflect how the perception of assessment reliability has evolved and how the documents stated it should be realised. The main changes could be summarised in the following six points.

- 1) In the 2009 and 2010 documents, only the GES assessment instruments (i.e., speaking and writing sections of the final test) were addressed in the standardisation policies. However, the standardisation policies released in 2011 addressed also the AES assessment instruments (i.e., report and presentation) (see row 2 of Appendix 1).
- 2) In the 2009 and 2010 documents, the policies included instructions about two processes (i.e., standardisation and moderation). In the 2011 document, the policies addressed three processes (i.e., standardisation, marking and moderation).
- 3) The meaning of the concept "moderation" seems to have changed across the 2009, 2010 and 2011 documents to be more about reconciling discrepancies in teachers' scores rather than analysing test items and scores across colleges (see rows 6 and 7 of the same appendix). In the 2009 and 2010 documents, post-moderation was stated to "be carried out by the Programme Director with regard to comparisons of scores between colleges and by the Assessment Team with regard to item analysis". However, in the 2011 document, post moderation was introduced as "discrepancies arising from individual biases are likely to be resolved through reference to a third party".
- 4) Both the 2009 and 2010 documents acknowledged the English language departments' failure to meet the set principles of standardisation and moderation (see row 1 of appendix 1). The 2011 document expected challenges in applying its policies (see row 4 of Appendix 1).
- 5) The 2009 and 2010 documents recommended standardising FP assessment by carrying out workshops where samples of written scripts and oral interviews were marked so teachers would have a feel of what the scores represented before marking the rest of the reports and interviews. The documents, however, did not specify the method of obtaining early samples of the reports and interviews. This point was raised in the 2011 document

where the policies advise conducting several presentations and collecting several scripts for standardisation and moderation purposes before commencing with marking all scripts and presentations (see row 2 of Appendix 1 for the 2009 and 2010 documents and rows 3 and 4 for the 2011 document).

6) Finally, the 2009 and 2010 documents dealt with the cross college standardisation as a comparison of students' scores in Language Knowledge quizzes and written assessment scripts amongst colleges. The 2011 document addressed the same issue more comprehensively where samples from presentations, reports and speaking tests were required too (see row 2 of Appendix 1 for the 2009 and 2010 documents and rows 3 and 4 for the 2011 document).

Regardless of the discussed process of adapting and refining a set of policies for moderating and standardising the FP assessment in and across the colleges, in practice, standardisation across colleges has been limited to the writing section of the GES tests only as has been affirmed by a member of the directing team (personal communication, April 1, 2012). CAS is still struggling to standardise marking the AES assessment tasks.

7.4 Replication of National Academic Standards in FP Specification

As the FP is expected to be audited in the near future, its documents (i.e., course specifications, FP handbook, assessment handbook ... etc.) intentionally and systematically were designed to adhere GFP standards to the letter. The intention to fully comply with these standards was stated in the *Foundation Programme 2010-2011* document.

The programme must meet the Oman Accreditation Council's General Foundation Programme Standards. These standards apply to all higher education institutions in Oman, private and public and compliance with the standards is **mandatory** by academic year 2010-11(CAS, 2010d, p. 1).

The GFP standards provided a set of learning outcomes that could guide HEIs to understand what was expected of a foundation programme. A comparison of these standards with the FP learning outcomes indicated that the standards seemed to be closely followed by FP course specifications, but there were real doubts about how closely (see Table 7.). The similarities and sometimes equivalence of FP and GFP's learning outcomes raises doubts about whether the process of writing the Foundation Programme learning outcomes involved any planning or consideration of the unique situation of the students at CAS.

These doubts were strengthened by the fact that the listening and speaking learning outcomes of the AES course were listed in the course specifications with a note saying that they were not covered by the textbooks and teachers should provide appropriate materials to meet them. Also, in the AES course, the students were not evaluated on the listening and reading skills which are part of the course specifications. This seems to suggest that the writing, reading, speaking and listening learning outcomes of the AES course were copied from the GFP standards as part of a blind matching process, possibly in order to perform well in the upcoming audition mentioned above.

Table 7. Similarities between AES learning outcomes and the GFP standards

		Document	
		CAS English Foundation Course Specifications (2010)	Oman Academic Standards for the General Foundation Programs (2008)
Skill	Reading	Read an extensive text of around 1000 words broadly relevant to an area of study and respond to questions that require analytical skills, e.g., prediction, deduction, inference (2010, p. 18)	Read an extensive text broadly relevant to the student's area of study (minimum three pages) and respond to questions that require analytical skills, e.g., prediction, deduction, inference (p. 10).
	Writing	Produce a written report of a minimum of 500 words showing evidence of research, note taking, review and revision of work, paraphrasing, summarising, use of quotations and use of references (p. 19).	Produce a written report of a minimum of 500 words showing evidence of research, note taking, review and revision of work, paraphrasing, summarising, use of quotations and use of references (p. 10).
	Listening	Take notes on longer talks/mini-lectures (10-15 minutes) (p. 19).	Take notes and respond to questions about the topic, main ideas, details and

		opinions or arguments from an extended listening text (e.g., lecture, news broadcast) (p. 10).
Speaking	Prepare and deliver a talk of at least 5 minutes. Use library resources in preparing the talk, speak clearly and confidently, make eye contact and use body language to support the delivery of ideas. Respond confidently to questions (p. 19).	Prepare and deliver a talk of at least 5 minutes. Use library resources in preparing the talk, speak clearly and confidently, make eye contact and use body language to support the delivery of ideas. Respond confidently to questions (p. 10).

The underlined phrases in the *CAS English Foundation Course Specifications* (2010) in Table 7 are identical to those in the *Oman Academic Standards for the General Foundation Programs* (2008), shown on the right hand side of the table.

It can be clearly seen from the table that the AES learning outcomes do not only address similar areas to those of the GFP standards, but are very comparable and identical in language. This finding might explain the mismatch between the focus of AES textbooks and that of AES learning outcomes, as has been mentioned previously.

8. Discussion

The four main issues raised above will now be discussed and linked to previous studies. These issues were: (1) conflicts and tensions in using norm and criterion-referencing principles, (2) compatibility between what is assessed and what is taught in AES and GES, (3) inconsistency in implementing assessment criteria, and (4) replication of the GFP standards in the FP specifications. These four areas could be considered as evidence on the content validity of FP assessment.

8.1 Norm vs. Criterion-Referencing Tests

Document analysis revealed that the stated intention of using criterion-referenced assessment in the FP was blurred by the actuality of using norm-referencing procedures in GES tests construction and analysis. Policy documents issued by OAAA and CAS clearly stated that assessment in the FP should be criterion-referenced not norm-referenced. Likewise, policy documents on the FP implied that criterion-referenced assessment was used, yet the GES test writing and analysing instructions in the same documents involved comparing students against each other, which is a characteristic of norm-referenced tests. Bachman (2004) says that aiming at most scores to be around the 50% mark of the test scores range is a characteristic of norm-referenced tests, in which the distribution of the scores should be normal, whilst criterion-referenced tests tend to be negatively skewed showing that most of the students have mastered the course objectives. In this study, it was found that the GES test writing and analysis procedures showed norm-referencing attributes implied in the stated test-writer instructions to compare the students against the low, medium and high groups of achievement. Also, the instructions dictated that the test items with difficulty indices of 0.25 or lower should be investigated for a positive correlation with the high achievers' scores. These procedures are clearly characteristics of norm-referenced tests (Bachman, 2004).

When a test is norm-referenced, mastering the learning outcomes does not become a priority. Consequently, some students can pass the FP without mastering all its stated learning outcomes. Thus, criterion-referenced assessment has been widely enforced by policy makers (Brindley, 2001; Lorena, 2007; Llosa, 2007). Sizmur and Sainsbury (1997, p. 129) refer the appeal of criterion-referenced assessment to the need to ensure the "minimal standards in basic skill areas, and the need to produce reliable measurement of these". In line with this view, the purpose of disseminating the GFPs document was stated to "seek to help ensure that those programs (GFPs) are effective in helping students attain the prescribed students learning outcomes" (2007, p. 4). Moreover, Sizmur and Sainsbury (1997) argue that criterion-referencing cannot be considered as a trait of a test; it is a concept that is defined by the interpretations made about the test scores and how they are used. If the test was designed to compare students performances against each other and the scores were analysed following the same purpose, then the used test makes norm-referenced interpretations of students English language abilities, thus the test shows attributes of norm-referencing. Applying this understanding to the context of this study, we can conclude that the GES test interpretations did not conform to the GFP standards when it made norm-referenced interpretations, however, the AES assessment tasks (i.e., report and presentation) made criterion-referenced interpretations.

In discussing the wash-back of English language tests, Shohamy (2007, p. 126) points out that “language policy documents often become no more than declarations of intent that can easily be manipulated and stand in stark contradiction as the ‘tested language’ obtains prestige and recognition”. A similar argument can be made about the use of norm-referenced tests when actually criterion-referenced tests were recommended by policy documents which were used as “declarations of intent”.

8.2 Incompatibility between What Is Assessed and What Is Taught

Several writers in the field of language testing argue that there should be a clear link between what is tested and taught in achievement tests (e.g., Bachman 1990, Fulcher & Davidson, 2007; Weir, 2005). Comparing the documents on assessment specifications, learning outcomes and content of textbooks revealed a clear incompatibility between what is taught and what is assessed. In both AES and GES courses, there were examples of how the intended course outcomes were matched by parallel test tasks, but underrepresented by the course materials. In the GES course both the writing and reading test tasks were at higher levels than the textbook tasks. In the AES course, the incompatibilities appeared in the writing scale used to mark the essays. The focus of the marking descriptors was substantially different from the writing learning outcomes. The descriptors highlighted the procedures of writing and submitting the essay more than the content and language accuracy of the essay. In the AES assessment, the incompatibilities also appeared in the speaking and listening learning outcomes mentioned in the course specification which were not covered by the textbook or assessment tasks. Though Hughes (2003) proposes that achievement tests should be built on stated objectives, not actual teaching, to generate positive wash-back effect, others (e.g., Weir, 2005) argue against this proposition and stress that achievement tests should be based on prior learning experiences not on intended ones. In the present context at least, Weir’s view is more pervasive

The above instances of incompatibility suggest a serious issue with the validity of FP assessment. Messick (1996) argues that there are two major threats to assessment validity which he entitles: construct underrepresentation, and construct-irrelevant difficulty. The criteria used in the AES essay marking scale, as shown by the results, underrepresented language accuracy and overemphasised procedures and technicalities of writing such as incorporating teacher comments or submitting on time. Incorporating teacher comments could be a very useful step in the process of writing but it should not be overstressed at the expense of other important language related criteria such as paraphrasing or using appropriate modal verbs and pronouns. Likewise, the GES test embodied features of construct-irrelevant difficulty in the listening task by testing students on an unfamiliar genre. Though some aspects of the AES tasks and GES test showed features of lower validity, it cannot be claimed that they were invalid assessment instruments. Messick advised that compelling evidence from multiple sources should be accumulated to evaluate assessment validity.

8.3 Inconsistency in Implementing Assessment Criteria

Though the policies of assessment standardisation and moderation were inaugurated in 2009, and were amended in 2010 and 2011, the process of implementing these policies still faced challenges in practice. The main two challenges were identified to be:

- 1) How scripts or recordings for the writing and speaking tasks could be obtained prior to the presentation or essay submission date for standardisation purposes in colleges;
- 2) How cross-college standardisation in marking the writing and speaking component of the assessment could be accomplished.

The *Assessment Policies* document (CAS, 2011) proposed that some of the presentations/speaking tests should be conducted in advance to be used as samples for marking the rest of the presentations. Also, it was suggested that a standardisation session should be conducted after the essays were submitted using a sample of the submitted scripts. All of these measures were intended to ensure consistency in marking the speaking and the writing components of assessment in the colleges, but they did not address cross-college standardisation. Also, the policies seemed to be suggestions more than commands. The results from analysing the policy document suggest that the moderation and standardisation policies were not all applied in practice.

Similar issues have been highlighted in the literature: Brindley (1998), in a review of studies on outcome based assessment, found that this type of assessment raised concerns about the validity of the descriptors and the objectivity of teachers’ judgements. He asserted that empirical studies showed instances of subjective and interpretation-based marking even when the scales were deemed to be clear by the teachers.

8.4 Replication of GFP Standards in FP Specifications

Language assessment in education has been affected by the international trend through ensuring accountability in

reporting achievement through using outcomes based assessment, as indicated earlier (Brindley, 2001; Llosa, 2007). Llosa (2011) explained that the rationale for standard based reforms was “to improve the quality of education for all students by developing rigorous standards and aligning instruction, assessment, professional development, and resources to those standards” (p. 367). Similarly, the FP in Oman is obliged to comply with GFP standards produced by the OAAA. The results of document analysis showed that the FP did not only (on paper) comply with the national standards; its AES learning outcomes actually replicated the ones in the GFP document. The GFP standards were used as the basis for the AES marking scales, not as guiding standards for what should be taught in classrooms. This finding can partly explain some of the students’ and teachers’ concerns about the difficulty levels of AES assessment.

9. Summary and Concluding Remarks

In this Study, the findings of thematic analysis of various types of documents were presented in four main headings. The first was how norm-referenced tests were used instead of the criterion-referenced tests mandated by the national and CAS policies on language assessment; it was argued that norm-referenced tests should not be used in FP assessment as they can have serious negative consequences. The chapter then explored inconsistencies amongst learning outcomes, materials taught, and assessment specifications; these inconsistencies were linked to the blind replication of the GFP standards. The third part revealed difficulties in standardising and moderating marking processes and highlighted inconsistencies in using marking scales, which will recur in the findings from other sources in the following chapters. The fourth investigated the language skills required in FY courses by analysing course specifications, required learning outcomes and actual test papers. This analysis concluded that the CS learning outcomes and assessment instruments, including the final test, seemed to rely on students’ language skills more than did the learning outcomes and assessment instruments of the other specialisations.

References

- Al Hajri, F. (in press). Student Perceptions of the Foundation Programme Assessment in Two Colleges in Oman. *Arab World English Journal*, 4(4).
- Al-Lamki, S. M. (1998). Barriers to Omanization in the private sector: The perceptions of Omani graduates. *The International Journal of Human Resources Mangement*, 9(2), 378-400. <http://dx.doi.org/10.1080/095851998341143>
- Ball, S. J. (1998). Big policies/small world: An introduction to international perspectives in education policy. *Comparative Education*, 34(2), 119-130. <http://dx.doi.org/10.1080/03050069828225>
- Atkinson, P., & Coffey, A. (2004). Analyzing documentary realities. In D. Silverman (Ed.), *Qualitative research* (pp. 77-90). London: Sage Publications.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511667350>
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing*, 15(1), 45-85. Retrieved from <http://ltj.sagepub.com/content/15/1/45.abstract>
- Brindley, G. (2001). Outcomes-based assessment in practice: some examples and emerging insights. *Language Testing*, 18, 393-407. Retrieved from <http://ltj.sagepub.com/content/18/4/393.abstract>
- Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. Retrieved from <http://www.emeraldinsight.com/journals.htm?articleid=17021410>
- Allwright, J., & Banerjee, J. (1997). *Investigating the accuracy of admissions criteria: A case study in a British university*. Institute for English Language Education, Lancaster University, Lancaster.
- CAS. (2009). *English Department Assessment Handbook*. Muscat: Ministry of Higher Education.
- CAS. (2010a). *Colleges of Applied Sciences Regulations*. Muscat: Ministry of Higher Education.
- CAS. (2010b). *Course Specifications for Foundation English*. Muscat: Ministry of Higher Education.
- CAS. (2010c). *English Department Handbook*. Muscat: Ministry of Higher Education.
- CAS. (2010d). *Foundation Program 2010-2011*. Muscat: Ministry of Higher Education.
- CAS. (2011). *Assessment Policies*. Muscat: Ministry of Higher Education.

- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272. <http://dx.doi.org/10.1017/S0267190599190135>
- Cotton, F., & Conrow, F. (1998). An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania. *IELTS Research Reports*, 1, 72-115. Retrieved from www.ielts.org/PDF/Vol1_Report4.pdf
- Donn, G., & Al Manthri, Y. (2010). *Globalization and Higher Education in the Arab Gulf States*. Oxford: Symposium Books.
- English Department. (2011). *Instructions for Report Writing and Presenting in AES Course*. CAS.
- Forster, N. (1994). The Analysis of Company Documentation. In C. Cassell, & G. Symon (Eds), *Qualitative Methods in Organizational Research*. London: Sage.
- Fox, J. (2004). Test decisions over time: Tracking validity. *Language Testing*, 21(4), 437-465. <http://dx.doi.org/10.1191/0265532204lt292oa>
- Fulcher, G., & Davidson, F. (2007). *Language Testing And Assessment: An advanced resource book*. London: Routledge.
- Hughes, A. (2003). *Testing English for Language Teachers*. Cambridge. UK: Cambridge University Press.
- OAAA. (2009). *Oman academic standards for general foundation programs*. Muscat: OAAA.
- Phillips, N., & Brown, J. L. (1993). Analyzing communication in and around organizations: A critical hermeneutic Approach. *Academy of Management Journal*, 36(6), 1547-1576. Retrieved from <http://www.jstor.org/discover/10.2307/256821?uid=3738784&uid=2129&uid=2&uid=70&uid=4&sid=21103252446643>
- Llosa, L. (2007). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing*, 24(4), 489-515. <http://dx.doi.org/10.1177/0265532207080770>
- Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language Testing*, 28(3), 367-382. Retrieved from <http://ltj.sagepub.com/content/28/3/367.short>
- Lorena, L. (2007). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing*, 24(4), 489-515. Retrieved from www.stanford.edu/dept/SUSE/.../Language%20Testing_Llosa_2007.pdf
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education*, 14(1), 117-130. <http://dx.doi.org/10.1080/09695940701272948>
- Sizmur, S., & Sainsbury, M. (1997). Criterion referencing and the meaning of national curriculum assessment. *British Journal of Educational Studies*, 45(2), 123-140. <http://dx.doi.org/10.1111/1467-8527.00041>
- Weir, C. (2005). *Language testing and validation—An evidence-based approach*. UK: Palgrave MacMillan.

Appendix

Changes in standardisation and moderation policies in official documents of the English Language Department from 2009 to 2011

Assessment Handbook: English Dept. (April 2009)	Assessment Handbook: English Dept. (March 2010)	Assessment Policies: English Dept. (October 2011)
Standardization	Standardization	Standardization
Standardization is more of a problem in continuous assessment than in exams because each college is responsible for producing its own assessments and because the lack of common delivery dates means that each one must be different. However, if Level Coordinators follow the format guidelines above and if a bank	In the past we have asked colleges to design their own continuous assessment items and then compared them for moderation purposes. Unfortunately this approach has presented severe reliability problems because of varied levels of challenge and it has also meant an excessive workload for coordinators. From	It is essential that all assessments of writing or speaking, in either examinations or continuous assessment such as the project presentations and reports currently in use at Year 1 and Year 2, be preceded by standardisation sessions. These sessions should aim to:

<p>of CA material can be built up over time to both provide exemplars and to serve as an item bank, it should be possible to ensure that students are presented with the same levels of challenge across the colleges (p. 4).</p>	<p>Spring 2010 we are adopting a policy in which each Level Coordinator will take responsibility for writing one component of the continuous assessment for their level (p. 5).</p>	<p>Clarify rating scales Facilitate the sensitive and consistent application of rating scales by teachers through practice assessments of samples of written or spoken performance and collective discussion (p. 2).</p>
<p>Samples in standardization Samples of written work and samples of language knowledge quiz performances should be gathered for monitoring by the Level Coordinator as a matter of routine. Each time a quiz is carried out, the Level Coordinator should request samples of marked scripts for comparison and a sample of these should be forwarded to PD [Programme Director] English. A reasonable sample to be forwarded to PD English would be 5 marked written assessments and 5 marked language knowledge assessments from each level from each college (p. 5).</p>	<p>Samples in standardization Samples of written work and samples of language knowledge quiz performances should be gathered for monitoring by the Level Coordinator as a matter of routine. Each time a quiz is carried out, the Level Coordinator should request samples of marked scripts for comparison and a sample of these should be forwarded to PD English. A reasonable sample to be forwarded to PD English would be 5 marked written assessments and 5 marked language knowledge assessments from each level from each college (p. 6).</p>	<p>Samples in standardization Ideally, standardisation sessions should be carried out prior to examinations or submission of reports or performance of presentations so that sufficient time can be afforded for adequate discussion of sample material and consensus achieved. For this to be possible, the English Dept needs to build up banks of standardisation material for all of the following: the placement test (written samples), the Challenge Test (Parts 2 & 3) (written and spoken samples) ENGL 3001, 4001, 5001, 6001 Mid-Term and Final Examinations (written and spoken samples) ENGL 1111, 1222, 2111, 2222-55 Final Examinations (written and spoken samples) (p. 2).</p>
<p>Assessor standardization for writing The Assessment Coordinator and the Coordinator for each level should run a standardisation session close to the exam period in which the following activities are carried out: Step 1: Reviewing/discussing the criteria. There are many ways in which this can be done. One way of focusing teachers on the meanings and differences between criteria and bands is to cut-up the criteria into single band/criterion segments and to get the teachers to reassemble them. Step 2: Independent marking of a single script, followed by comparing of marks and discussion, followed by presentation of actual marks (as pre-determined by Level Coordinators and Assessment Team). Step 3: Marking of other scripts in the same way. A record should be kept of the marks each assessor gives for the last two scripts marked in the session.</p>	<p>Assessor standardization for writing The Assessment Coordinator and the Coordinator for each level should run a standardisation session close to the exam period in which the following activities are carried out: Step 1: Reviewing/discussing the criteria. There are many ways in which this can be done. One way of focusing teachers on the meanings and differences between criteria and bands is to cut-up the criteria into single band/criterion segments and to get the teachers to reassemble them. Step 2: Independent marking of a single script, followed by comparing of marks and discussion, followed by presentation of actual marks (as pre-determined by Level Coordinators and Assessment Team). Step 3: Marking of other scripts in the same way. A record should be kept of the marks each assessor gives for the last two scripts marked in the session. Assessors who are off by the</p>	<p>Assessor standardization for writing An alternative procedure for standardisation of writing may be considered where no body of old samples exists, which is this: immediately after submission of the project reports or assignments or collection of the examination scripts, the LC should take a sample of the material roughly assessed as fail, weak pass, fair pass, strong pass and present these to a meeting of the assessors. The scripts should then be assessed using the rating scales and consensus achieved as to appropriate marks. These marks should then be used as standards for subsequent marking of the remainder of the scripts or reports. This is a fair way of achieving standardisation within a college. However, it does not address possible cross-college differences so should not be used if samples of cross-college assessed material are available for pre-examination or pre-submission standardisation (p. 2).</p>

Assessors who are off by the end of a training session will need monitoring (p. 15).	end of a training session will need monitoring (p. 11).	
<p>Assessor standardisation for speaking</p> <p>Similar procedures should be carried out at similar times for speaking assessors.</p> <p>These sessions should involve:</p> <p>Step 1: Reviewing the criteria.</p> <p>Step 2: Reviewing the tests</p> <p>Step 3: Viewing, independent assessing, and discussion of a single recorded performance prior to discussion of actual marks (as pre-determined by Level Coordinators and Assessment Team).</p> <p>Step 4: Viewing of other recordings in the same way (p. 15).</p>	<p>Assessor standardisation for speaking</p> <p>Similar procedures should be carried out at similar times for speaking assessors.</p> <p>These sessions should involve:</p> <p>Step 1: Reviewing the criteria.</p> <p>Step 2: Reviewing the tests</p> <p>Step 3: Viewing, independent assessing, and discussion of a single recorded performance prior to discussion of actual marks (as pre-determined by Level Coordinators and Assessment Team).</p> <p>Step 4: Viewing of other recordings in the same way (p. 11).</p>	<p>Assessor standardisation for speaking</p> <p>Standardisation of speaking tests and presentations is much more difficult. If no cross-college assessed material exists, one policy that might be followed is for an LC [Level Coordinator] to schedule tests and presentations so that a very limited number may be carried out in phase 1 and recorded for standardisation purposes, permitting a discussion of the performances and the setting of standards for subsequent assessment of the remainder of the tests or presentations in phase 2 (p. 2).</p>
<p>Post-moderation of exams</p> <p>Post-moderation will normally be carried out by the Programme Director with regard to comparisons of marks between colleges and by the Assessment Team with regard to item analysis. However there are some important post-moderation tasks to be undertaken in the colleges as well. These are best done by Assessment Coordinator and/ or Level Coordinators (p. 19).</p>	<p>Post-moderation of exams</p> <p>Post-moderation will normally be carried out by the Programme Director with regard to comparisons of marks between colleges and by the Assessment Team with regard to item analysis. However there are some important post-moderation tasks to be undertaken in the colleges as well. These are best done by Assessment Coordinator and/ or Level Coordinators</p>	<p>Moderation</p> <p>Where principles 1-4 [under the heading “Marking”] can be maintained there should be little need for much post-assessment moderation. Discrepancies arising from individual biases are likely to be resolved through reference to a third party (HoD, Assessment Coordinator, Level Coordinator) during the marking process. Where moderation is important, is in those situations where 1-4 cannot all be maintained strictly, as in the case of project presentations. In such cases, HoDs, ACS or LCs should follow these steps:</p> <ol style="list-style-type: none"> 1. Take averages of each teacher’s scores. 2. Where unexpected differences occur, check with the teacher concerned to provide clarification. We cannot exclude the possibility that some classes are more able than others and some teachers are more able than others. There are perfectly reasonable grounds why differences may occur between classes working at the same levels, and moderation is not to be used to bring an artificial uniformity to test scores. 3. Where no such satisfactory clarification occurs, teachers’ marks

		should be brought into line with the average for the level. (p. 4)
No matching categories	No matching categories	<p>Marking</p> <p>It is impossible to eliminate individual biases in marking completely, even through the use of rigorous standardisation procedures. It is essential therefore that marking for exams and continuous assessments be organised with the following principles in mind:</p> <ol style="list-style-type: none"> 1. Wherever possible all marking of writing and speaking should be carried out by two people. 2. Wherever possible the class teacher of the students concerned should not be one of those two people. 3. Wherever possible the two markers should assess independently of each other (i.e., 'blind'). 4. There must be a third person to whom any differences in marking may be referred. <p>It is likely that these principles may be maintained in some circumstances but not all (p. 3).</p>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).