

Student Perceptions of the Foundation Programme Assessment in Two Colleges in Oman

Fatma Al Hajri

Ministry of Higher Education, Oman

Abstract

This study investigates student perceptions of the English language assessment in terms of its effectiveness, reliability, validity and the students' preference of its constituting instruments. It included 184 students from two Colleges in Oman who participated in responding to a questionnaire, of those 106 students participated in gender specific focus groups. The data obtained from the questionnaire was analysed for descriptive statistics and significant differences amongst the groups using Mann-Whitney U test and Kruskal Wallis Test. Student views expressed in focus groups were analysed using thematic content analysis. The results suggested that the students perceived FP assessment as generally effective but they were ill-informed about its nature and contents. The data also indicated that there was a perceived need for more assessment instruments and an unexpected low social impact of FP assessment considering its high stakes.

Keywords: Assessment, English Language, Higher Education, Tests, Continuous Assessment.

Introduction

In Omani higher education, a level of proficiency in English is a requirement to access most higher education institutions (HEIs), and English is considered a vital tool to access the national labour market (Al-Lamki, 1998, 2006; Donn & Al-Manthri, 2010; Al-Issa, 2006). Al Shemli (2009) looks at the role of English in higher education in the globalised context, and argues that “the main effect of globalisation in the Sultanate of Oman is the need to diversify the economy and raise standards; and the concomitant pressure to supply skilled graduates for rapidly changing economic conditions” (p.10). In this context, improving the English language skills of students is identified as a major challenge in higher education, though reforms have been undertaken at both the school and university levels (Al Shemli, 2009). Reforms in school education that target improving proficiency in English language alongside skills in other subjects are highlighted by Alsarimi (2001), who calls for innovative methods to assess these skills and lessen the use of assessment tools that solely rely on memorising or rote learning.

The new educational system aims to strengthen student competencies in mathematics and science, to improve student proficiency in English, and to teach students to use scientific methods and problem solving ... it is crucial that student assessment be reformed as well (pp. 27-28).

Changing the assessment instruments does not result in improving students' English language proficiency per se. How these instruments are implemented and how students and teachers view these instruments participate in their effective implementation. This paper investigates the student perceptions of FP assessment and suggests a number of areas to enhance its effectiveness.

Literature Review

As graduates' proficiency in English language is required by both the national and international labour markets, it has been identified as a vital asset in higher education. Though the internationality of English language as a lingua franca has also been emphasised as one of the reasons for this (Al-Issa, 2006; Al-Mahrooqi, 2012), the fact that the private labour market mainly operates in English has been seen by others as a more compelling reason. The need for graduates with an acceptable level of proficiency in English is clear in Al-Lamki's exploration of the barriers to Omanisation (i.e., replacing expatriates by Omani nationals in the labour market).

Since English is the international language of communication and is also the medium for international business transaction, and since English is the operational language in Oman's private sector, it is recommended that the level and standard of English taught in schools and colleges be improved (2011, p.395).

In response, the governing bodies responsible for education in Oman have set conforming goals. The Ministry of Education states that:

The government recognises that the facility in English is important in the new global economy. (Reform and Development of General Education, Ministry of Education, 1995, p. A 5-1: as cited in Al-Issa, 2006).

The Ministry of Higher Education proclaim similar views on the role of English language in CAS. English language teaching is associated with national development in Oman; the National English language Policy/Plan (NELP) states that:

the English language skills of Omani nationals must be seen as an important resource for the country's continued development. It is this recognition of the importance of **English as a resource for national development and the means of wider communication within the international community that** provides the rationale for English in the curriculum (Al-Issa, 2005 , p.2, emphasis in original).

Despite such stated intentions, plans and policies to promote the English language proficiency of the labour force, recent studies of graduates' English skills have found that these are inadequate for the needs of the private sector (Al-Mahrooqi, 2012; Al-Lamki, 2006). Al-Mahrooqi asserts that "research and experience have proved that the majority of school and college graduates possess neither adequate English language skills nor communication skills to function effectively in the workplace, which is dominated by expatriates from around the world" (2012, p. 124).

A similar view has been reported by the graduates themselves who "felt that their communication skills were very poor. Even the students on the verge of graduation expressed this, with much regret and sorrow" (Al-Mahrooqi, p.129). The students' consciousness of their lack of adequate language skills seems to have deterred them from applying for vacancies in the private sector; Al-Lamki reports that "students felt that the private sector discourages and disqualifies Omanis from applying because of the requirements for work experience and English language skills" (2006, p.392). She found that 72% of the 58 graduate students, in this study, considered English language a barrier to work in the private sector. All in all, the issue of employability is very complex; one can only speculate that factors such as motivation, proficiency in English language or possessing other skills might be relevant, but the magnitude of these roles is still under-researched.

A common concern that is raised about pre-sessional programmes in general is that they allow students to embark on academic study with an inadequate level of English proficiency (Allwright & Banerjee, 1997; Fox, 2004). Cotton and Conrow (1998) report that students expressed a need for extra EAP instruction even after they had reached the IELTS level required by their universities. Though most internationally recognised higher education institutions do not permit embarking on higher education before reaching a certain minimum level of English language proficiency, some others do allow students to start academic studies at lower levels of language proficiency and provide them with language support programmes (Fox, 2004). The Colleges of Applied Sciences (CAS) follow this approach: students are provided with two EAP courses in their first year and two English for Specific Purposes courses in their second (one each semester) to help them overcome some language challenges they might face when starting academic study. The role of assessment in enhancing language programmes and improving student proficiency has been highlighted (e.g., Alsarimi, 2001; Al kharusi, 2008).

This study was conducted in two Colleges of Applied Sciences: Sur and Rustaq. It is estimated that almost 80% of the students admitted to higher education in Oman require English language courses in the Foundation Programme (FP) before starting their 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). The aim of teaching English language is stated to be “equip[ing] students with both the English Language and academic study skills they will need to succeed in their subject studies” (CAS, 2010a, p.33).

The FP consists of twenty hours per week of English language instruction, and two hours of mathematics and/or computer skills courses in each semester. The English language programme is divided into two major courses, the General English language (GES) and Academic English Skills (AES), the acronym (FP) will be used to refer to the English language components only of the programme.

The present assessment system in CAS uses both standardised tests and Continuous Assessment (CA) as the way forward for education and assessment reforms in Oman. In a description of the assessment used in CAS, it is stated that “academic regulations mandate the allocation of 50% of marks to a final examination and 50% to CA” (CAS, 2010a, p. 35). Assessment in FP also included tests and CA instruments. Table 1 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 (i.e., CA).

Table 1. Assessment instruments in the foundation programme courses

Course	Assessment Instruments	% of Course Total	% of Foundation Programme Total
General English Skills	Mid-term Test	40%	50%
	Final Test	60%	
Academic English Skills	Presentation	50%	50%
	Report	50%	

In order to pass, students must obtain 50% of the total marks in each course; failing to achieve this means failing the FP and consequently being denied access to higher education.

Study Questions

This study investigates student views of FP assessment in terms of its effectiveness, reliability and validity, social and educational impact, and their preference of various assessment tools. It provides a preliminary, small case study that was used as part of a larger study conducted to fulfill the requirements of a PhD at the University of Edinburgh. This paper mainly looks into the following questions and triangulated data from two different data collection methods: a Questionnaire and focus group.

1. How effective was the process of assessing students' English language performance, through continuous assessment and tests, in the Foundation Programme (FP) as viewed by students?
2. How was the reliability and validity of FP assessment viewed by students and teachers?
3. How was the impact of FP assessment perceived by students and teachers?
4. 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 students?

Methodology

The Questionnaire

A total of 220 FP students were invited to participate in this study, which was conducted over two academic semesters. Of those a total of 184 students participated in responding to the questionnaire; 127 (69%) of them were from Rustaq College and the other 57 (31%) students were from Sur College. The sample consisted of 119 female students (64.7%) and 65 male students (35.3%). At CAS, when the students were admitted to the FP, they had already selected their intended specializations. The participants were from four different departments: Information Technology (IT), Communications Studies (CS), International Business Administration (IBA) and English Language-Education (see Table 2).

Table 2. The distribution of participants by specializations

Specialization	<i>n</i>	%
Information Technology	50	27.2
Communication Studies	25	13.6
International Business Administration	85	46.72
English Language-Education	24	13.0
Total (<i>N</i>)	184	100

The 184 students who completed the questionnaire forms were invited to take part in the focus groups and 106 of them agreed to participate; 12 focus groups were conducted seven of which were female only groups and the other five were male only groups. The participants expressed their preference of gender specific focus groups.

Table 3. An overview of the participants in focus groups

Group	College	Gender	Number of Students	Length/minutes
Group 1	Rustaq	F	12	53 min.
Group 2	Rustaq	F	8	32 min.
Group 3	Rustaq	F	16	32 min.
Group 4	Rustaq	F	9	35 min.
Group 5	Sur	M	9	33 min.
Group 6	Rustaq	F	6	32 min.
Group 7	Sur	F	12	38 min.
Group 8	Rustaq	F	13	26 min
Group 9	Sur	M	8	38 min
Group 10	Sur	M	3	14 min.
Group 11	Sur	M	7	34 min.
Group 12	Rustaq	M	13	51 min.
Total			106	418 min.

All of the groups' discussions were carried out in Arabic and were video-taped. The recordings were translated into and transcribed in English.

Data Analysis

Questionnaire

Appendix 1 displays the number and percentage of the students who responded to each item by selecting a point in the five-point likert scale questionnaire. The points 1 to 5 respectively denote Strongly Agree (SA), Agree, (A), No Opinion (NO), Disagree (D), and Strongly Disagree (SD). The appendix also shows the means of the students' responses to the items including the recoded ones.

It is important to note here that the views of the students and teachers on assessment validity are actually a form of "face validity", they are not evaluations of the "construct" or "content" aspects of assessment validity.

The questionnaire items were organised into groups for later analysis based on their themes. In some cases, shown in the table, this entailed recoding (i.e., changing responses to be 1=5, 2=4, 3=3, 4=2 and 5=1) so that semantically opposite or near-opposite items could be more directly compared. The expectation was not that the items in a group would be found almost totally equivalent, but that they might reveal broad trends of satisfaction, dissatisfaction and perhaps other feelings and perceptions, whilst anomalous response patterns might offer further insights.

These expectations were only partially met. In the large section on *Perceived Validity*, mean scores were fairly similar within and even between sub-sections: all were below 2.9 (3.0 = *No Opinion*), indicating broad but not overwhelming acceptance of FP assessment validity. In some other sections, however, means varied more widely and fell on both sides of the middle point (M=3.0). Some of these cases will be discussed below.

Focus Group

Though content analysis is sometimes linked to quantifying the elements of the content according to a set of categories in a systematic manner (Bryman, 2008), thematic content analysis is linked to qualitative data analysis. According to Bryman interpretation of Althiede's analysis (1996), ethnographic content analysis involved an element of "constant discovery and constant comparison of relevant situations, settings, styles, images, meanings, and nuances" (p.393). In this study, thematic (ethnographic) content analysis which focuses on "what is said rather than on how it is said" (Bryman, 2008, p.412) was used to analyze the transcripts of the student focus groups.

The term "coding" though is widely used; it usually entails different procedures that sometime authors do not explicitly describe (Richards & Morse, 2007). It is essential for the enhancement of the quality and validity of any study to delineate not only the procedures followed in data collection but also in data analysis (Creswell, 2011; Maxwell, 1992; Mishler, 1990). In this study, the transcripts were coded using topic coding which links the ideas to the data rather than labelling the data only (Richards & Morse, 2007). A list of 20 codes emerged from reading the transcripts and referring to the study questions. The codes were selected based on their re-occurrence of the ideas and relativity to the study questions. Once the list of codes was refined, the steps below, which were adapted from (Miles & Huberman, 1994), were applied to focus group scripts. These steps were:

- a. Assigning codes to the appropriate extracts in all interview scripts or audio recordings.
- b. Reading the extracts linked to each code and clustering them into groups.
- c. Looking for possible themes.
- d. Comparing and contrasting the themes within the same phase and between the phases.

- e. Splitting or combining themes.
- f. Building a logical chain of evidence.
- g. Making conceptual coherence.

The codes were categorised into three main themes: uncertainty about assessment instruments' weightings and scales, tests in students' perceptions, and CA in students' perceptions. The latter two themes were divided onto six subthemes that covered the students' views on the reliability and validity of FP assessment.

Results

Students' Responses to the Questionnaire

This section aggregates the average responses to the individual items and presents the Mean and Standard Deviation of the responses to each topic to obtain an overview of students' perceptions of *Perceived Reliability*, *Perceived Validity*, *Preference of Tests*, *Satisfaction with Current Assessment Practices*, and *Impact of FP assessment*. Table 4 lists the questionnaire's topics which are hierarchically ordered according the means of the responses to each topic.

Table 4. Means of the student questionnaire's topics

Topics	Mean	Std. Deviation
Political Impact	1.81	.78
Preference of Continuous Assessment	1.99	.75
Perceived Construct Validity	2.06	.75
Preference of Tests	2.36	.79
Perceived Content Validity	2.73	1.04
Perceived Reliability	2.83	.78
Social Impact	2.85	.58
Satisfaction with Current Assessment Practices	3.17	.51

The results showed that the students seemed to positively perceive the reliability and validity of the FP assessment as the mean scores were *Perceived Reliability* (M =2.83), *Perceived Construct Validity* (M=2.06) and *Perceived Content Validity* (M=2.73). It can be seen from appendix 1 that the students seemed to show less satisfaction with the content validity of FP assessment than they did with its construct validity. A closer look at the elements of the *Content Validity* topic reveals that the mean of one of its items was close to the disagreement range (i.e., $M \geq 3.1$). The means for the four items were respectively item 1.1: M = 2.7, item 1.2: M = 2.4, item 1.3: M = 2.9, and item1.4: M = 2.3 (appendix 1). The students' responses to the third item implied that their certainty level of how their achievement would be exactly assessed in the FP courses was not high. Actually, 41% of the students responded with *Disagree* or *Strongly Disagree* to this item, while 47% of the students responded with *Agree* or *Strongly Agree*. It seems that a considerable percentage of the students were ill-informed about how they would be assessed in their English language courses. The students' lack of knowledge about the assessment procedures could have lowered the average mean of FP *Perceived Content Validity*.

Another interesting point in Table 4 is that the mean score of the perceived *Political Impact* of FP was lower than that of perceived *Social Impact* of FP assessment. The means of the responses to FP *Political Impact* and *Social Impact* topics are 1.8 and 2.85 respectively, both of which fall in the agreement range (from M= 1 to M= 2.9). There seemed to be a majority agreement with the statements that indicates that FP assessment could entail considerable political impact by affecting the job opportunities in the labour market and the country's international status. Also, there seemed to be a moderate agreement with the idea that assessment in FP did not entail negative or drastic social impact on students' lives. Though FP assessment could be considered high-stakes, most of the students felt that the assessment was relatively fair, not frightening, and did not depend on luck. It is worth noting that though most of the students felt that both Continuous Assessment (CA) and tests were not stressful, they seemed to believe that the tests (item 6.1: M= 2.26) were less stressful than the CA (item 6.2: M=2.1). Some of the reasons for this view were discussed in focus groups.

The topic of *Satisfaction with Current Assessment Practices* had the second highest mean (M= 3.17). This implies that most of the students were not satisfied with the FP assessment. Investigating the items under this topic shows that though most of the students seemed generally satisfied with the FP assessment (item 5.1: M=2.15), most of them also believed that the FP assessment should be changed (item 5.2: M= 3.42) and that the change should not include fewer assessment instruments (item 5.2: M=2.29). Interestingly, this response was found to conform to what most of the students said in the focus groups about increasing the number of assessment instruments. In general, it could be concluded that the students seemed to be satisfied with the assessment practices, but they tended to believe that there should be more assessment instruments.

The last point about the means of the responses to questionnaire topics is that the respondents seemed to prefer AES continuous assessment (item 4.2: M= 2.23) more than the GES tests (item 3.1: M=2.46). This preference resonated with their opinions as expressed in the focus groups as discussed below.

Comparing Perceptions amongst the Groups

This section further explores the students' responses to the questionnaire by investigating significant differences in responses to the items of each topic amongst the groupings by gender, college, specialization and self-evaluation. This exploration aims at identifying any clear pattern of consistent differences in the groups' responses that might shed some light on the participants' perceptions using filters such as college or gender.

Before looking at the results of these tests it is important to clarify the rationale for using non-parametric tests. First of all, the data generated by a likert scale is an ordinal and sometimes categorical type of data that is best investigated using non-parametric techniques (Pallant, 2007). Second, these two tests were selected because the data set was found to be not normally distributed, so non-parametric tests are ideal in this situation (Fielding & Gilbert, 2006). The data set of students' responses was tested for normality of distribution using Kolmogorov-Smirnov test, skewness values, histograms and box plots. The results showed that the distributions of students' responses to each topic violated the assumptions of a normal distribution. Thus, Mann-Whitney U Test was used to investigate significant differences between two groups and Kruskal Wallis Test was used amongst three groups or more.

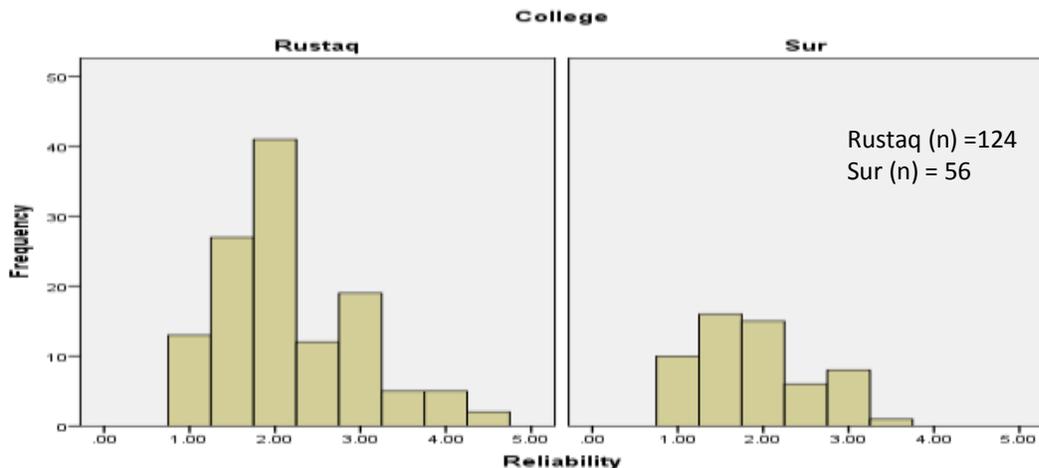
Differences between College Groups

In the college groups, a significant difference was found between Sur College’s students (Mean Rank= 95.49, n= 56) and Rustaq College’s students (Mean Rank= 79.45, n= 124) in their responses to *Perceived Reliability*, $U= 2,853$, $p<.05$. This indicates that the students at Sur College were significantly different in their perception of *Perceived Reliability* than the students at Rustaq College (see Figure 6.1). Sur Students viewed the assessment tools as more reliable than did their peers in Rustaq College. This can be deduced from the Mean Rank values presented above and Mean values in Table 5.

Table 5. Means of students’ responses to questionnaire topics by colleges

Topics	Rustaq College’s Students		Sur College’s Students	
	Mean	n	Mean	n
Social Impact	2.38	96	2.26	55
Preference of CA	1.93	125	2.01	55
Preference of Test	2.31	126	2.45	56
Political Impact	1.79	124	1.80	57
Perceived Reliability	2.16	124	1.90	56
Perceived Construct Validity	2.14	122	2.02	55
Perceived Content Validity	2.70	121	2.38	57
Satisfaction with Current Assessment Practices	2.83	114	2.71	52

Figure1. Students’ responses to perceived reliability by colleges



Differences between Gender Groups

The results showed that the male and female students differed in responding to two topics namely *Preference of CA* and *Political Impact*. The Mann-Whitney U test revealed a significant difference between the female students’ responses (Mean Rank=80.61, n=116) and male students’ responses (Mean Rank= 106.13, n=62), $U= 9,351$, $Z= -3.2$, $p=0.001$. The female students showed more *Preference of CA* than did the male students. Likewise, the male students’

responses (Mean Rank=102.52, n=62) and female students' responses (Mean Rank=84.18, n=118) were significantly different on the *Political Impact* of FP assessment: the female students seemed to believe that FP assessment had higher *Political Impact* more than did the male students, $U= 2912$, $Z=-2.3$, $p=0.21$. All in all, this means that the female students seemed to prefer CA more than did the male students and they seemed to emphasise the political impact of FP more than did the male students (see Table 6 & Figures 2 &3).

Table 6. Means of students responses to questionnaire topics with gender

Topics	Male Students		Female Students	
	Mean	n	Mean	n
Social Impact	2.32	58	2.34	92
Preference of CA	2.23	62	1.80	116
Preference of Test	2.45	62	2.30	118
Political Impact	2.00	62	1.69	118
Perceived Reliability	2.13	62	2.04	117
Perceived Construct Validity	2.09	61	2.11	115
Perceived Content Validity	2.57	61	2.59	115
Satisfaction with Current FP Assessment Practices	2.73	51	2.83	105

Figure 2. Responses to preference of CA by gender

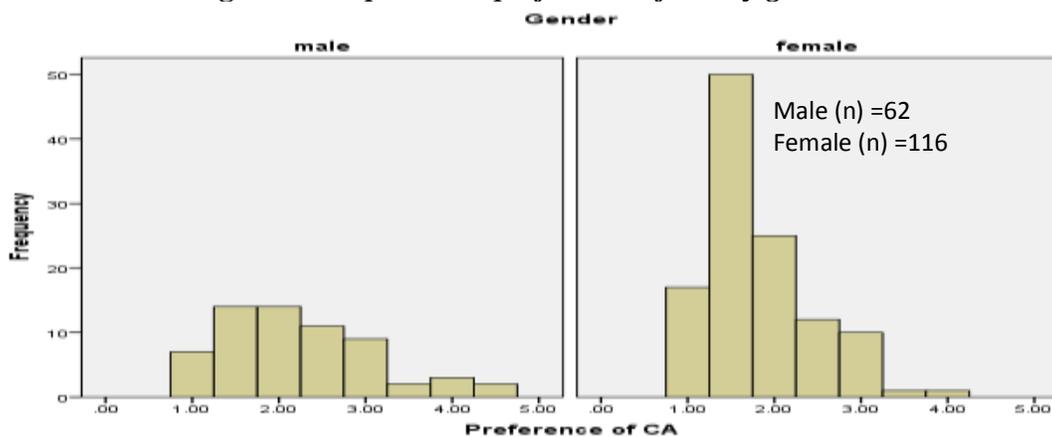
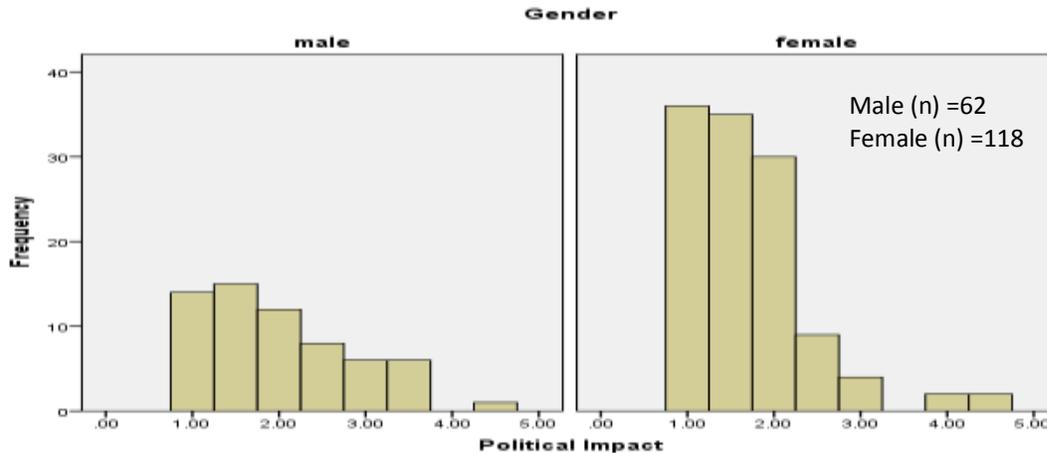


Figure 3. Responses to political impact by gender



However, the difference between the female and male students' Preference of CA was not matched by a significant difference in their AES continuous assessment scores as shown by the results obtained from using Mann-Whitney U test. Actually, the female and male students' mean grades in both of the FP assessment instruments (i.e., AES continuous assessment and GES tests) showed no significant differences (see Table 7).

Table 7. Mean and quartile of scores in GES and AES courses by gender

Gender		GES	AES	
Male	n	Valid	61	
		Missing	3	
	Mean	1.82	2.71	
	Quartile	25	1.30	1.70
		50	1.70	3.00
		75	2.30	3.70
Female	n	Valid	101	
		Missing	18	
	Mean	1.95	2.85	
	Quartile	25	1.70	2.70
		50	2.00	3.00
		75	2.30	3.30

Differences among Self-Evaluation and Specialization Groups

In the first section of the questionnaire the students were asked to self-evaluate their language proficiency by selecting one of five levels to represent their language proficiency levels (i.e., 1= poor, 2=average, 3= good, 4=good, and 5= excellent). Kruskal Wallis was used to explore the differences, and the results showed no significant differences in the students' responses to the questionnaire topics with regard to their self-evaluation groups. Similarly there were no significant differences in responding to the questionnaire's topics among the specialization groups.

To summarise, it could be concluded that the responses of participants to the questionnaire's items were not significantly different between the specialization and self-evaluation groups, but they were between the college and gender groups. The college groups showed a significant difference in responding to *Assessment Reliability*, and the gender groups showed a significant difference in responding to both *Preference of CA* and *Political Impact* topics.

Focus Groups

This section presents the results attained from the student focus groups which are intended to partly answer the questions of the study stated above.

Uncertainties about GES and AES Assessment Instruments

The majority of the students in seven focus groups tended to express uncertainty about how the scores were distributed on the main FP assessment instruments. Though, all of them seemed to be aware that the assessment of the General English Skills (GES) course included a midterm test and a final test and the assessment of the Academic English skills (AES) course included essay writing and a presentation, many of the students seemed uncertain about the scores distribution. The following extracts reveal this uncertainty as demonstrated in the students' discussion of the weightings of the assessment instruments. This extract comes from Focus Group 9.

Student 3: In the GES course, there will be 50% of the total scores on the presentation and 50% on the essay.

Student 7: It is still not very clear how the scores in the AES course are divided. Some teachers say the project is allocated 50% of the total scores while other teachers say that it is worth 20% of it only, so we do not know yet...

Student 3: It is not clear. In the GES course, we have a speaking test and in the AES course, we have a presentation. So are they accumulated and how many scores each is awarded?

In addition to the uncertainties about scores distributions, many students seemed ill-informed about the criteria used in marking scales to evaluate their language performances in the essay and presentation. When asked about how scores would be given, most of them were aware that their teachers would be using marking scales but seemed oblivious of the scales' criteria. Few students, in three groups only, mentioned several criteria of the scales such as: eye-contact, posture and grammar with regard to the presentation marking scale; and grammar, organization and content with regard to the essay marking scale. The following extract manifests the lack of clarity of the marking scales as experienced by many students.

Group (3)

Student 1: This semester the way we are going to be assessed is not clear. No one explained to us. The course plan seems not stable and the teachers seem not sure of how and what the assessment will look like

Students' uncertainties were not limited to the distribution of scores and marking scale criteria, there were also uncertainties about the test sections. In one group, few students stated their confusion about whether grammar would be included in the midterm and final tests or not. In another group, the students' discussion suggested that they were not aware of the fact that grammatical rules were actually tested in a section in the test titled "Language Knowledge". They kept on speculating whether the grammar rules would be tested or not.

GES Tests in Students' Perceptions

The presentation of students' views of FP assessment categorises the views into two main sections: views about GES tests and views about AES assessment. In each of the categories, examples of students' discussions were provided.

The Content of GES Tests

In the focus groups, issues about the content of the midterm test and forthcoming final test were raised and debated. One of the issues was the difficulty of the reading tasks in which they faced new topics. One instance mentioned was that in the midterm test the reading test task was about "human cannonball". They said it was a new type of sport that they had never heard of before, thus they found it difficult to respond to the task questions. Group (11) discussed this issue saying:

Student 3: The reading passage was incomprehensible without the picture.

Student 6: True, it was about an unknown kind of sport.

Student 1: We have never heard of such a thing so comprehending what the passage was about, was so difficult.

It was discussed that in the midterm test, it was not only the topics that were new to them, but also the vocabulary of the reading task and length of the writing task.

The second issue raised was about the lack of proper preparation for the grammar test tasks. Though the midterm and the final tests allocate only 10% of the total mark to the Language Knowledge task (i.e., grammar and vocabulary test items), almost in all focus groups, the students expressed their need for additional grammar tutorials. In several groups, it was reported that even though the textbooks included activities on grammar rules, the teachers tended not to teach them. This was because they, as many students believed, seemed to be unqualified to teach grammar rules or because the teachers expected the students to study the rules by themselves as a form of autonomous learning. The students felt that more grammar lessons were needed to succeed in their future academic study and to pass the GES tests as manifested throughout the following extracts.

Group (9)

Student 1: We did the midterm exam and it was very difficult. We had not been given any practice quizzes before it. It was a shock.

Student 2: Our teachers do not explain grammar and we found the grammar part of the test very difficult. None of the teachers discuss grammar with the students.

The listening task, on the other hand, seemed to be considered by almost all groups as the most difficult task of the GES tests. Likewise, it was claimed that the listening activities undertaken in classroom were simpler than what was in the tests and fewer than what was needed to be able to perform well in the listening test tasks. The difficulty of the listening test task was conveyed in the extracts below.

Group (11)

Student 2: The listening (part two) was so difficult. We are not used to such a thing. We need a book on listening to practice listening.

Student 5: It was very quick; we could not answer the questions in the pace that we were supposed to.

The Consequences of GES tests

When asked about the fairness of the assessment instruments and importance of passing the FP assessment, the students' responses varied from arguing that the assessment was very fair and passing was very important to claiming that assessment was unjust and passing FP assessment was unimportant. In almost all of the focus groups, FP assessment seemed to be regarded as unfair because of the distribution of scores on test tasks, type of test tasks, or inappropriate curriculum. It seems that the meaning of the concept "fairness" did not only include tests' qualities but was stretched to include course curricula.

Group (3)

Student 12: Tests are not fair; they test grammar more than the other skills. Most of the scores are on grammar and since we are weak in grammar we lose a lot of scores in the tests.

Student 13: I felt that the test let me down. I was depressed because of my low scores.

The significance of passing the FP assessment was considered differently in the focus groups. Most of the students believed that they would definitely pass the FP and refused to consider the possibility of failing. For few students, considering the consequences of failing triggered negative social and psychological connotations such as: shame and depression.

Students' Perceptions of AES Continuous Assessment

Like the previous section, this section presents the results of the student focus groups on AES assessment categorised into sections. Categorizing the evidence from the focus groups is intended to facilitate understanding the students' views on FP assessment effectiveness. These categories were not used to imply a view of distinct types of validities rather they were used to present validity evidence on a clearer way.

What AES Continuous Assessment Measures

Though the report and presentation were considered as good assessment instruments *per se*, many students seemed to believe that CA was not suitable for everybody and it did not fully reflect their language skills. Two main reasons for this belief recurred in the focus groups. The first reason was that writing and presenting could result in a performance inhibition caused by students' personal traits (e.g., low confidence and shyness) or learning styles (e.g., auditory, visual, and kinaesthetic). The second reason was the lack of proper guidance, training and practice on writing and presenting (e.g., different criteria sets used by different teachers). The intertwining of the students' opposite feelings of appreciation of CA's role in FP assessment and worry of its shortcomings are apparent in the following discussion.

Group (1)

Student 3: But if you were a silent student by nature, presentations and other oral means might not be just in terms of assessing students levels.

Student 8: I think it should be looked at as a comprehensive thing, I mean assessment. The tests with the presentations complement each other in terms of assessing students English language levels and showing their abilities. Some students are more capable of undertaking the exams while the others are more capable of presenting so the various ways of assessing the students give a fair chance to all.

Student 5: Still presentations influences confidence and does not show the real level.

The Content of AES Continuous Assessment

Few students in three groups doubted the content of the AES assessment (i.e., essay and presentation) and said that the essay was too complicated for their English language levels and research skills; and that they sometimes intentionally plagiarised or cheated in other ways. These discussions went as follows:

Group (7)

Student 12: They [the teachers] teach us steps on how to write an essay but never ask us to practice. We need to practice in class or out but the teacher always says that he will not mark our work. How do they know that the end result is my own work if they do not see samples throughout the semester? Last semester I asked my sister to write the essay for me and I will do the same this semester because I simply cannot write it though I know the steps.

Consistency in Implementing AES Marking Scales

As has been pointed out earlier, most of the students seemed to be aware of the fact that the marking scales will be used AES assessment but not of how they will be used. It was apparent from most focus groups that many students felt that the marking scales were inconsistent in how they were implemented or interpreted.

Group (7)

Student 12: But teachers differ in terms of the criteria they use to assess the students. We know that other groups are told different things about how they will be assessed in the essay. This is wrong, we are not assessed equally. All students should be given the criteria at the beginning of the semester before starting to write the essay or preparing for the presentation.

The Feedback Given in Continuous Assessment

A recurring theme in focus groups was the lack of teacher feedback offered in the essay and presentation. Generally in most of the groups the students expressed dissatisfaction with the amount and nature of feedback provided; and argued that the appropriate feedback could improve their language skills. It was claimed that sometimes the feedback imparted was ambiguous, negative, delayed or non-existent. The subsequent extracts display some of the students' comments on teacher feedback.

Group (8)

Student 2: there is no feedback at all, we do not see our scores and we do not know how we are doing so far, we just wait until the end of the semester and wait to see the result at the end. We should have been given some feedback to lead us on what we should be learning and how we should do things right.

Student 6: we need more quizzes more things to tell us about our levels and guide us in learning. But most importantly we need feedback on the tools.

The Consequences of Continuous Assessment

The essay and presentation were described as subjective and unfair by most of the groups. Teaching styles, marking scales, scores distribution and availability of resources were all considered factors that participated in characterising CA as being unfair.

Group (8)

Student3: The fairness of assessment depends on the teachers. Some teachers are unfair in marking the exams. For example sometimes we cannot revise the exam results with the

teacher or discuss any concerns some of them get angry when you try to discuss the scores with them.

Discussion

FP Assessment Effectiveness in Student Perceptions

Most students seemed less satisfied with the GES assessment (i.e., tests) than they were with the AES (i.e., presentation and report). The content of the GES tests was severely criticised by students. The students emphasised the difficulty that they faced in the grammar, reading and listening sections of the mid-term test. They elaborated that the reading topic was new; the grammatical rules were not all covered in the course; and the listening genre had not been introduced to them before.

Though AES assessment was generally positively viewed by the students, they both made comments signalling its problematic content and construct. They raised three concerns about the essay: (1) high difficulty level, (2) plagiarism, and (3) variability in implementing marking scales. In focus groups, some students admitted to committing plagiarism because they found the essay very difficult for them to write using their own words. Students' interactions with the assessment tasks have been identified as a parameter in understanding students' performances and difficulty of assessment tasks (Bachman, 2002). Hamilton (2003) discussed a number of studies that investigated cheating in tests; one study (Jacob and Levitt, 2003) found that the cheating instances increased when the tests were high-stakes. Another study on students' perceptions of plagiarism in higher education found that students sometimes perceived plagiarism as "a strategy for coping with the demands of higher education level work and the pressure to succeed" (Ashworth, et al., 1997, p. 194). A similar perception was documented in other studies in the field of second language learning and assessment (Currie, 1998; Pecorari, 2003). The findings of this study conform to the findings of the studies that have recognised task difficulty as a contributing factor that influences students performances (Bachman, 2002); and considered it a principal factor in resorting to plagiarism (Hamilton, 2003).

Furthermore, the difficulty level of the essay task was not the only element of AES assessment criticised; the students expressed their apprehension of the inconsistency in implementing the marking criteria. This concern seems to match similar concerns documented in several studies on performance assessment (Brindley, 1998, 2001; Hay & Macdonald, 2008). Brindley (1998) reviewed a number of articles and identified numerous problematic issues with the validity of the scales used to mark students' performances; he categorised them into political, technical and practical. He asserted that "subjective judgements of language performance are likely to show a good deal of variability" (p.65). Addressing this concern, Gipps (1999) advised that rater inconsistencies should be minimised to reach a better reliability especially in high-stakes assessment tasks. Given the high-stakes nature of FP English language assessment, and the concerns raised by the students about inconsistency in implementing marking measures, there seems to be an urgent need for implementing the standardization and moderation procedures discussed.

Uncertainties about the FP Assessment Elements

In the focus groups, most students seemed uncertain about specific aspects of the AES and GES assessment instruments. Some of these aspects were: the weightings of the assessment

instruments and the criteria of the marking scales. Empirical evidence have suggested that the students' understanding of assessment requirements might well be different to that of their teachers' as Green (2007) indicated in reference to Weir and Green's study (2002). In line with this suggestion, this study found that indeed students expressed a less certain understanding of what was required by the assessment activities than that of their teachers'. In AES assessment, for instance, most students complained about the lack of information on some aspects of which their teachers seemed very well aware.

Students' uncertainties about aspects of FP assessment could be also referred to the unavailability of sample or mock tests. It was reported that past exams were not accessible for them and consequently they were not completely aware of the exams' structure and contents. Rea-Dickins (1997) asserted that in centralised systems where teachers were not involved in assessment development, they could be not "prepared sufficiently for the task of implementations" (p.308). In the context of the current study, though GES tests were written by individual assessment coordinators from different colleges, the tests were not distributed to the rest of the teachers several of whom were novice in the Colleges and had never seen these tests before. Understandably, several teachers and many students seemed ill-informed about FP assessment.

Perceived Need for More Assessment Instruments

Regardless of the previously mentioned concerns about the effectiveness of FP assessment, students tended to express the need for more assessment instruments. The students' declared a request for more assessment tasks can indicate and be linked to the need for more feedback. Administering additional assessment tasks and feedback might appear unrelated but they actually are when considering the findings of the studies conducted on feedback suggesting that summative assessment provides less feedback than formative assessment. Brindley (1998), in a comparison of summative and formative assessment, stated that the former was more suitable for the purposes of policy makers and educational bureaucrats for its skimmed aggregated details, while formative assessment provided detailed and elaborated feedback. Broadfoot (2007) identified the purpose of summative assessment as to "sum up to progress of an individual in relation to some given criterion" (p.110), and the purpose of formative assessment as to provide "information to be used as feedback to modify the teaching and learning activities they are engaged in" (p.111). In the higher education context, there seems to be a move towards less formative assessment and more summative assessment with late or insufficient feedback (York, 2003). Revisiting the findings of this study, it could be noticed that both GES and AES assessment instruments might be considered as summative with regard to the time and type of feedback with which the students are provided. Though the students received some sort of feedback on the first and second drafts of the essay; this feedback, as considered by the students, was occasionally detrimental, late or insufficient. Even, the few students who seemed to believe that feedback on the essay was appropriate attributed their satisfaction of the feedback provided to having a good teacher.

Hamilton (2003) reviewed a number of studies that provided evidence of better students' performance when more feedback on how to improve performance was given to them. Likewise, reviews on the effectiveness of feedback showed that it varied based on different aspects (Bangert-Drowns, et al., 1991). Recent reviews that focused on what the feedback was about (i.e., task, processing, regulatory) found that feedback was most effective when it attended self-

regulation (Hattie & Timperley, 2007). A conforming finding was reported by Black (2003, as cited in Broadfoot, 2007), who asserted that ‘task-oriented’ feedback enhanced the ‘learning power’ of the students and enabled them to take control of, and encouraged them to get involved in their own learning.

Tests vs. CA in Student Perceptions

Most of the students preferred CA over the tests according to their questionnaire’s responses, although, both were considered useful. These results differ from those of a recent study by Cheng, et al. (2011) which explored students’ and parents’ perceptions on the traditional examinations compared to a currently applied School Based Assessment (SBA) system in China. They found that no significant difference in how students viewed SBA and exams. They also reported that the students differed in how they perceived the SBA and exams items based on their self-reported language levels; “students with high perceived language competence responded more positively to the items relating to the external examinations while students with low perceived language competence responded more positively to the items relating to SBA” (p. 238). The results of the current study did not report any significant difference amongst students’ self-evaluations of their language proficiency levels in their perception of CA compared to tests. Actually all self-evaluation groups responded more positively to CA than they did to the tests.

Similar to the results obtained from the questionnaires, the students’ views in focus groups seemed to generally prefer CA more than the tests for several reasons. The students’ preference of CA seemed to be propelled by their appreciation of the process of learning that takes place in the evaluation process of the students’ language proficiency. This apparent students’ appreciation of learning through assessment is in line with and reinforces the voices calling for “assessment for learning” as a way forward in assessment for its ability to improve students’ performances as supported by empirical evidence (Broadfoot, 2007).

Student Views of FP Assessment Impact

Interestingly, students’ responses to the *Social and Political Impact* topics in the questionnaire were similar in rating the political impact higher than the social impact and in implying that FP assessment entailed political but not serious social consequences on students, teachers and society. They seemed to recognise the “prestige” and importance of the English language assessment for future national employment and international status of the country, as well as its role as a gatekeeper to higher education; this finding conforms to previous studies (e.g., Shohamy, 1996; Ross, 2008). However, the students disagreed with the view that FP assessment had drastic social consequences. The social impact was not considered great in the students’ responses. Their responses showed a majority agreement with the items that suggested that FP assessment was fair, not frightening and not stressful.

This finding is substantiated by the findings obtained from the focus groups. An unexpected result obtained from the focus groups was the relatively moderate to non-existent social impact of FP assessment considering its high-stakes nature. Failing in FP could mean that students lose their scholarships to study at CAS or become suspended for one academic year during which an accredited proof of a specific language level should be attained from a recognised private language teaching institution. However, not only very few students expressed that failing in FP assessment could entail a negative social stigma, most of them seemed to be confident that they would pass and did not show any concern of failing in FP. Shohamy (2001) explained that there

are a number of factors that could contribute to understanding the consequences of a test like language status, purpose of assessment, format of assessment and low/high-stake nature of tests. Though all of these factors when considered in the context of FP assessment predict a strong negative social impact, the findings of this study arrived at a different conclusion. A possible logical explanation for this finding is what the students indicated about inconsistent implementation of FP assessment marking criteria.

When the students were asked whether the tests and CA were not stressful in the questionnaire, their responses were positive. But they agreed more with the statement “tests are not stressful” than with the statement “CA is not stressful”. This finding is in line with the argument that performance based tasks involve communication stress or anxiety which may well influence students’ performance along with other factors (Bachman, 2002; Philip, 2011). It was also found that performance assessment did not produce better results than test in terms of the writing skill (Hamp-Lyons, 1997); therefore, if AES assessment, which includes performance based tasks, against common expectations (e.g., of teachers), did not provide a less stressful environment than tests, and did not result in better performance, the advantages of using this type of assessment in FP assessment should be reviewed.

The results showed that the students seemed to be very confident of passing the FP assessment which indicates a low educational impact of FP assessment. However, the students revealed their concern that their language level would be lower than what is required for FY Study. The FP results in 2011 showed that more than 90% of the students in both colleges passed to the first academic year, though, the teachers generally expected less than 80% of their students to pass (as another part of the PhD study suggested). In Sur College 92% of the FP students successfully passed and in Rustaq College 97% of the FP students passed. The students’ and teachers’ view of FP assessment inability to fulfil the role of filtering the linguistically able students to study in FY, could pose a threat to the validity of FP assessment.

Conclusion

This study investigated students’ views of FP assessment. There are some limitations and recommendations that can be deduced from the results obtained. One limitation is using focus groups in this study as it posed some challenges. These challenges included training participants in this method of sharing views, re-coding all opinions, transcribing and analysing generated data. The other two main challenges were the “no shows” which is a common and documented disadvantage of focus groups (Bryman, 2004), and ensuring free and equal expression of views.

There are several recommendations that can be inferred from the findings, two of them are discussed here. First, there are uncertainties about the assessment content and structure. Assessment details should be shared with both the students and teachers at the beginning of the academic semester to eliminate any underperformance due to uncertainties and increase the validity of FP assessment. Second, the students shared that they received a lack of feedback that could be attributed to the summative nature of FP assessment. Instead, formative assessment instruments that provide enough feedback to students and that show a high degree of validity should be considered. This could be achieved by incorporating smaller units of classroom assessment early in the semester to allow enough time for feedback. These units should be validated prior to use and teachers should be trained to mark them as consistently as possible, preferably using a similar marking scale to that used for other performance assessment tasks.

About the Author:

Fatma Al Hajri is a lecturer at the Colleges of Applied Sciences – Sur. She attained her PhD from the University of Edinburgh in Language Assessment and Programme Evaluation. She did her MA in Applied Linguistics at the University of Queensland. She worked as an English language lecturer for several years before being appointed as the Deputy Director for the English Language Programme at the Colleges of Applied Sciences in the Ministry of Higher Education, Oman.

References

- Al-Issa, A. (2005). The implications of the teacher educator's ideological role for the English language teaching system in Oman. *Teaching Education*, 16(4), 337-348.
- Al-Issa, A. (2006). The cultural and economic politics of English language teaching in the Sultanate of Oman. *Asian EFL Journal Quarterly*, 8 (1) 194-218.
- Al Kharusi, H. (2008). Effects of classroom assessment practices on students' achievement goals. *Educational Assessment*, 13, 243-266.
- Allwright, J and Banerjee, J, 1997, *Investigating the accuracy of admissions criteria: A case study in a British university*, Institute for English Language Education, Lancaster University, Lancaster.
- Al-Lamki, S. M. (1998). Barriers to Omanization in the private sector: The perceptions of Omani graduates. *The International Journal of Human Resources Management*, 9(2), 378-400.
- Al-Lamki, S. M. (2002). Higher education in the Sultanate of Oman: The challenge of access, equity and privatization. *Journal of Higher Education Policy and Management*, 24(1), 76-86.
- Al-Lamki, S. M. (2006). The development of private higher education in Oman. *International Journal of Private Education*, 1, 54-77.
- Al-Mahrooqi, R. (2012). English communication skills: How are they taught at schools and universities in Oman? *English Language Teaching*, 5(4), 124-130.
- Al Shemli, S. H. (2009, September). Higher education in the Sultanate of Oman: Planning in the context of globalization. Paper presented at *IIEP Policy Forum*. UNESCO: Paris.
- Alsarimi, A. M. (2001). New trends in assessment in the Sultanate of Oman: Goals and characteristics. *Educational Measurement: Issues and Practice*, 27-29.
- Ashworth, P., Bannister, P., & Thorne, P. (1997). Guilty in whose eyes? University students' perceptions of cheating and plagiarism in academic work and assessment. *Studies in Higher Education*, 22(2), 187-203.
- Altheide, D. L. (1996). *Qualitative Media Analysis*. Thousand Oaks, Calif: Sage.
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453-476.
- Bangert-Drowns, R. L., Kulik, C.-L. C., & Kulik, J. A. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213-238.
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing*, 15(1), 45-85.
- Broadfoot, P. (2007). *An Introduction to Assessment*. New York: Continuum.
- Bryman, A. (2008). *Advances in mixed-methods research: Theories and applications*. London: Sage Publications.
- Black, P. (2003). Assessment for Learning- Lessons from Research and from Practice. Paper presented in *Beyond Exams conference*, Bristol.
- 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.
- Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221-249.
- 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
- Creswell, J. W. (2011). Controversies in mixed method research. In N. Denzin, & Y. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (pp. 269-283). London: SAGE.

- Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing*, 7(1), 1-18.
- Donn, G., & Al Manthri, Y. (2010). *Globalization and Higher Education in the Arab Gulf States*. Oxford: Symposium Books.
- Fielding, J., & Gilbert, N. (2006). *Understanding Social Statistics*. London: Sage Publications.
- Fox, J. (2004). Test decisions over time: Tracking validity. *Language Testing*, 21(4), 437-465.
- Fox, J. (2008). Alternative assessment. In E. Shohamy, & N. H. Hornberger, *Encyclopedia of Language and Education: Language Testing and Assessment* (pp. 97-109). New York: Springer.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355-392.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university prepositional language courses. *Assessment in Education: Principles, Policy and Practice*, 14(1), 75-97.
- Hamilton, L. (2003). Assessment as a policy tool. *Review of Research in Education*, 27, 25-68.
- Hay, P., & Macdonald, D. (2008). (Mis)appropriations of criteria and standards-referenced assessment in a performance based subject. *Assessment in Education*, 15(2), 153-168.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14, 295-302.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Jacob, B., & Levitt, S. (2002, August). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*. 118(3), pp. 843-77
- Mishler, E.G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review*. 60 (4). 415-443.
- Maxwell, J. A. (1992). Understanding and Validity in Qualitative Research. *Harvard Educational Review*. 62 (3). pp. 279-301.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 24-1256.
- Miles, M. B., & Huberman, A. M. (1994). Making good sense. In M. B. Matthew (Ed.), *Qualitative data analysis: An expanded sourcebook* (pp. 245-287). London: Sage.
- OAAA. (2009). *Oman academic standards for general foundation programs*. Muscat: OAAA.
- Pallant, J. (2007). *SPSS survival manual*. New York: Open University Press.
- Philip, E. M. (2011). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal*, 76(1), 14-26.
- Richards, L., & Morse, J. M. (2007). Coding. In L. Richards (Ed.), *Read me first for a user's guide to qualitative methods* (pp. 133-151). London: Sage.
- Rea-Dickins, P. (2007). Classroom-based assessment: Possibilities and pitfalls. In J. Cummins, & C. Davidson, *International Handbook of English Language Teaching* (pp. 505-520). Springer.
- Ross, S. J. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing*, 25(1), 5-13.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13, 298-317.
- Shohamy, E. (2001). *The power of tests*. Harlow: Longman.
- Weir, C. J. & Green, A. (2002) *The impact of IELTS on the preparation classroom: stakeholder attitudes and practices as a response to test task demands*. Unpublished IELTS Research Report.(The British Council).
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.