



Perceptions of the selection criteria of Omani English language teachers: Implications for policy reconsideration

Ali Al-Issa^{a*}, Ali Al-Bulushi^b, Rima Al-Zadjali^c

^a*Sultan Qaboos University, P.O. Box 50, Al-Khodh 123, Sultanate of Oman*

^b*Sultan Qaboos University, P.O. Box 50, Al-Khodh 123, Sultanate of Oman*

^c*Sultan Qaboos University, P.O. Box 50, Al-Khodh 123, Sultanate of Oman*

APA Citation:

Al-Issa, A., Al-Bulushi, A., & Al-Zadjali, R. (2017). Perceptions of the selection criteria of Omani English language teachers: Implications for policy reconsideration. *Journal of Language and Linguistic Studies*, 13(2), 235-259.

Submission Date: 9/4/2017

Acceptance Date: 29/7/2017

Abstract

Proficiency in the English language has been described as central for determining Non-Native English Speaking Teachers (N-NESTs) selection for joining the profession. The Ministry of Education in the Sultanate of Oman decided to set the score of Band 6 on the International English Language Testing System (IELTS) for accepting the English Language Teaching (ELT) graduates of Sultan Qaboos University (SQU) as full-time teachers. This qualitative content analysis study hence triangulates data from interviews and the relevant literature about the different language proficiency measurement devices to determine the selection of SQU graduates to join the ELT force. The findings show that the Ministry of Education needs to rethink the decision of opting for the IELTS as a valid and reliable selection criterion. The results have important implications for setting selection criteria of becoming an English language teacher in the Sultanate of Oman and beyond.

© 2017JLLS and the Authors - Published by JLLS.

Keywords: English language teaching; student teachers; IELTS

1. Introduction

1.1. English and English Language Teaching Today

The English language today is unquestionably a global phenomenon. Its importance as a global lingua franca is undisputed. It is a commodity that is shared by the international community and no single individual or organization can claim its ownership. English at present is the dominant language of academia, business, science and technology (Roux, 2014). Clyne (2008) describes English today as being “indispensable” and perceives that it “. . . will continue to be such in the foreseeable future, for both survival and communication across many nations and cultures” (p. 10). This situation has driven many governments worldwide to revise their ELT policies. On the top of these policies has been the selection and hiring of English language teachers and the criteria that determine such a decision.

*Corresponding author. Tel.: +968-99320225
E-mail address: ali2465@squ.edu.om

The introduction of the Communicative Language Teaching (CLT) approach has thus brought new perspectives and challenges about English Language Teaching (ELT). One of these perspectives and challenges has been the role of teachers as proficient language users with adequate knowledge to influence and facilitate their students' language learning and acquisition and help prepare them for an ever changing and challenging world and demanding job market (Al-Issa & Al-Bulushi, 2012).

1.2. English and ELT in Oman

English has been accepted as the only official and preferred international language and a fundamental tool for national development and modernization since the Sultan took up the reins of the Sultanate of Oman in 1970 (Al-Issa, 2014a). English in Oman has institutionalized domains like the media, education, and business (Al-Busaidi, 1995), and it is learnt for different significant purposes, which justifies its choice by the Omani government. The Omani government has invested heavily in ELT through introducing the demanding and challenging CLT in the school ELT curriculum through the revolutionary multi-million dollar Basic Education System (BES) innovation since 1997 (Al-Issa & Al-Bulushi, 2012).

1.3. Statement of the Problem

Several authors (Klanrit & Sroinam, 2012; Nkata, 2010; Tajeddin & Adeh, 2016) have emphasized the pivotal role of English language proficiency for Non-Native English Speaking Teachers (N-NESTs). They further highlighted its implications for such teachers' confidence and performance inside and outside the ELT classroom and for their students' language learning.

Thus, a large number of students graduating every year from high school in Oman have failed to demonstrate the required adequacy in English, which has been partly attributed to the teachers' linguistic incompetence (Al-Issa, 2005; Al-Issa, Al-Bulushi, & Al-Zadjali, 2016; Moates, 2006). The Omani Ministry of Education, therefore, decided to recognize the International English Language Testing System (IELTS) as the official benchmark to assess the ELT Sultan Qaboos University (SQU) graduates' English language proficiency. It is noteworthy that SQU was established in 1986 and is the only state-owned university in Oman. The Ministry has approved an overall Band 6 as a minimum requirement to be eligible to apply for an ELT job at its schools to help facilitate the implementation of the linguistically demanding CLT-oriented BES innovation. Furthermore, graduating teachers are expected to prepare more linguistically adequate students with skills and ability to meet the increasing challenges of globalization and internationalization.

However, the choice of the IELTS as an English language proficiency indicator has generated a substantial debate in the Omani ELT community due to different validity and reliability issues related to the test. Furthermore, the quality of the College of Education ELT program came under the microscope, as its credibility and reliability were questioned. It is worth mentioning that SQU has been pursuing accreditation and international recognition through the National Council for Accreditation of Teacher Education (NCATE) and the American Council on the Teaching of Foreign Languages (ACTFL) for its different programs for the past two decades or so to improve its world ranking. It has hence invested generously in this regard. Unlike the College of Medicine and College of Engineering, which were accredited, the College of Education, for example, has been making great efforts to achieve this target and has succeeded in achieving international recognition of some in its programs.

1.4. Literature Review

IELTS is an internationally recognized test which provides a benchmark for making comparisons. It broadly measures whether candidates are ready to study or train in the medium of English. It covers the four language skills, which are equally weighted. Focus in the IELTS is more on demonstrating high-order thinking skills and strategies in listening, reading, speaking, writing, and vocabulary using a wide range of materials and task and text types and forms. This kind of performance measurement should help “achieve a more accurate understanding of language proficiency in relation to real-world target domains” (Biber, Gray, & Staples, 2014, p. 4).

Hinton (2009) too acknowledged the comprehensiveness of the IELTS as a test with high face validity. However, he argued against the poor criterion-related validity and content validity and recommended conducting further research about the IELTS by external bodies to address these areas. Other authors like Veerappan and Sulaiman (2012) and Karim and Haq (2014) questioned the inter-rater reliability of the IELTS in writing and speaking respectively. Hinton (2009) further viewed the IELTS as having a negative washback at the “micro” level (Pierce, 1992) due to the large number of discrete-point item types, which lack communicative behaviour authenticity of the four skills being learnt (Messick, 1994, p. 241) and allow for too much guessing. Hinton argued that this caused negative impact on teaching and learning.

Alshammari (2016) evaluated the IELTS academic reading module in detail. He examined its validity and reliability and discussed how far its format, operations, conditions, and techniques met its purpose. The author found that the module considered the necessary academic reading skills that university students were likely to perform. Nonetheless, he recommended revising the strict test scoring rules regarding spelling, grammar, and the number of words required for written responses to improve the test scoring validity in the light of its proposed purpose.

Dooley and Oliver (2002) found that the IELTS reading module had the highest predictive validity of academic success for the business students but not for the science or engineering students. However, the two authors stressed that the IELTS “did not give any guarantee of success” (p. 52), as a host of other contextual factors beyond the receiving institution’s control influenced the respective candidates’ language proficiency. Such findings brought the issue of validity into question.

Moreover, Roshan (2013) critically reviewed the latest version of the revised IELTS speaking test introduced in 2001 and found that it was more valid than the Test of English as a Foreign Language (TOEFL) in the sense that candidates were given planning and preparation time in the second phase of the test, which impacted communication conceptualization. It also allowed candidates to engage in more natural interactive communication situations with less predictable questions, which largely represented real world situations. Furthermore, the rating system was changed from holistic to analytic, which enhanced validity. Roshan found that the IELTS was criticized for being subjective, which led to unfairness in the testing process. Other weaknesses were pertinent to the number of examiners (one) and the time devoted to testing speaking (15 minutes), which had negative implications for the test validity.

Within the speaking test vein, Iwashita, Brown, McNamara, and O’Hagan (2008) argued against the validity of the academic IELTS and problematized the accuracy of the assessment levels it provided in the speaking skill. They also questioned the relationship between the different aspects of performance and the overall judgment of a candidate’s proficiency. Iwashita et al. (2008) additionally argued for the complexity of second language and second language acquisition. Moreover, the authors argued against the development and measurement of the IELTS to include knowledge beyond grammar and vocabulary. They concluded their work by recommending rethinking scale development in particular and test preparation and teaching/learning in general.

Griffiths (2003) critically commented on the IELTS and TOEFL's validity and stated that neither the "expensive and time consuming" (p. 38) TOEFL nor the IELTS "is universally accepted as a valid measurement of proficiency" (p. 37). He held that while the former lacked construct validity, some components of the latter were marked subjectively like the writing and speaking sections.

Freimuth (2016) also investigated the construct, predictive, and content validity of the IELTS through analyzing the content of 24 IELTS Task One in order to determine whether Emarati (from the United Arab Emirates – UAE, a neighbouring country of Oman and shares borders with it) university students anecdotal claims of cultural bias in on the IELTS academic writing exam were valid. She stressed the importance of suitability of the exam to all candidates and that "if one cultural group has more familiarity with the test item (task) or its topic (content), then it is said to house cultural bias which impacts the exam's validity as a whole, but in particular its content validity" (p. 2). The author found that the cultural bias perceived by the Emarati students had some validity, given the socio-cultural and educational background of the students, which was largely similar to the Omani students'. She hence concluded that the IELTS did not fit the purpose it was written for and that it was unfair, unreliable, and invalid to some candidates with different needs like the ones investigated in her study and in this study.

A different dimension of the IELTS's cultural bias was also found in Freimuth's (2013) investigation of 60 reading passages for cultural capital. The author found that the readings referred to 139 places or regions around the world with only five references pertaining to the Middle East and none to the UAE, where she conducted her study. The author argued that this kind of cultural bias not only disadvantages students in terms of background knowledge, but it also slows them down – "both of which can have significant impact on their reading scores" (p. 18). Freimuth concluded that using the culturally biased IELTS as a one-size-fits-all instrument for measuring language proficiency on a mass scale should be avoided.

An additional study, which investigated the predictive validity of the IELTS test scores as a graduation requirement for a teacher education program in the UAE was carried out by Khemakhem (2016) on 27 subjects. He used four different IELTS speaking tests from different published resources that had been used by the examiners for the mock test. The researcher found that the use of the IELTS cut-off score (Band 6) as a predictor of post graduation performance in schools may not be justifiable based on differences in context-related interaction and their impact on language proficiency. He further found that there was no significant relationship between lexical diversity scores and between the IELTS speaking test scores and the subjects' grades on teaching practice. The author concluded that the use of the IELTS' score for high stake decisions as in the case of the bachelor degree student teachers at the Higher Colleges of Technology in the UAE was "wrong".

Another study, which questioned the IELTS's construct validity, was conducted by Al-Malki (2014). She collected 94 graduate freshmen Omani English teachers' IELTS, GPA (Grade Point Average), and teaching competencies. The author found a moderate significant relationship between the IELTS and the participants' GPA. However, the relationship between the IELTS and teaching competencies was weak. Al-Malki recommended that the Omani Ministry of Education reconsidered its overall Band 6 requirements on the IELTS for appointing the College of Applied Sciences graduates as full-time English teachers. She further recommended that the Ministry allowed those who scored below Band 6 on the IELTS to teach since they had been learning at the college for 5 years to become teachers.

Interestingly, some researchers (e.g. Dooley & Oliver, 2002) found no link between the IELTS scores and academic performance. On the contrary, other researchers (e.g. Arcuino, 2013; Avdi, 2011; Feast, 2002; Kerstjens & Nery, 2000; Woodrow, 2006; Yen & Kuzma, 2009) found generally positive

(although sometimes weak or inconsistent) correlations between the IELTS entry levels and GPAs. Bayliss and Ingram (2006) found that the IELTS could predict their 28 international students' language behavior enrolled in different courses at University of Melbourne, despite the fact that those students perceived their language proficiency levels quite differently. Arcuino (2013) suggested that external and cultural factors played a part in the student's academic success when the relationship between the IELTS score and final GPA was weak. The author hence concluded stated that a high score in this international test was not a true reflection of one's academic success.

On the subject of academic performance and teaching competencies as fundamental aspects leading to scoring a high GPA, Freeman, Katz, Gomez, and Burns (2015) thus proposed an "English-for-Teaching" construct. The construct reconceptualized English language proficiency of teachers to help them teach in public sector schools. The construct additionally highlighted the essential English language skills teachers need to prepare and deliver the lesson in a national curriculum in English in a way that is recognizable and understandable to their non-native English speaking students. The authors were critical of establishing a connection between improving teachers' English proficiency and improved instructional quality and student learning, which is the case within the context of this study.

Freeman et al. (2015) argued for teachers developing a sense of authority and expertise as a result of their knowledge of the "local" tasks and responsibilities of their teaching situations and the social and interactional contexts of their classrooms. They suggested that this should be represented in "managing the classroom," "understanding and communicating lesson content," and "assessing students and giving them feedback" (p. 134) while using a global language like English, for example.

Freeman et al. (2015) discussed several implications drawn from the English-for-Teaching construct. The first was pertinent to designing teacher education programs that addressed "teachers' language development in terms of tangible, classroom-based outcomes" (p. 137). The authors acknowledged that this enhances program face validity and facilitates better use of teachers' time and energy. Second, Freeman et al. argued for the construct as creating assessments linked to teachers' classroom proficiency thus allowing for stronger validity. Last but not least, the authors discussed implications for enunciating criteria for developing teachers' language and professional knowledge. They argued that this potentially provided information about how local performance was aligned with global set benchmarks.

Cranston (2012) thus urged administrators to opt for valid and reliable selection criteria that genuinely reflect teaching effectiveness and success. Jacob, Rockoff, Taylor, Lindy, and Rosen (2016) thus argued for GPA and interviews as strong predictors of teacher effectiveness, but found a weak link between them and receiving a job offer or being hired. Cranston problematized and questioned the administrators' current hiring practices and the extent to which they led to hiring strong and capable teachers. He also held that hiring teachers is a complex process that requires administrators to consider several criteria prior to making a recruitment decision.

In his exploratory study, Cranston (2012) hence developed and administered a 23-item questionnaire on 28 superintendents in public school divisions' central offices in a Canadian province to solicit their feedback about the processes and criteria, and their respective weightings used in their hiring practices. Findings indicated that interviews were ranked highest (100%) for making a decision to offer a teaching job amongst the eight specified criteria.

In her review of current studies and directions for future research on employment interviews, Macan (2009) acknowledged that applicants always expected interviews to be a fundamental part of an employment selection process due to their fairness. Macan further argued for "structured" interviews as displaying high levels of parallel-forms reliability and inter-rater reliability and criterion-related,

construct validity and content validity, depending on their design, implementation, the purpose they are used for, and job complexity.

1.5. Purpose and Rationale

No studies supporting the implications of the IELTS's, interviews, and classroom observation reliability and validity to ELT policy implementation have been found in the pertinent literature. Besides, a lack of collecting data through qualitative means has been criticized for being another weakness. These are gaps that ELT research at SQU has suffered from (Al-Issa, 2015), and which this study attempts to compensate for.

Moreover, this study has important implications for similar contexts around the world. It is expected to add to the wealth of studies that examined the reliability and validity of the IELTS as an international language proficiency yardstick.

1.6. Research Questions

We are motivated to ask the following research questions:

1. What are the STs' perceptions about the validity of the IELTS?
2. What are the STs' perceptions about the reliability of the IELTS?
3. What are the STs' perceptions about the validity and reliability of interviews and classroom observation as alternative selection criteria?
4. What implications do the STs' perceptions about the reliability and validity of the alternative selection criteria have for English language teachers' policy formulation?

2. Method

2.1. Design

In this exploratory research study we use a qualitative content analysis approach, in order to gain an understanding of the STs' underlying perceptions about the reliability and validity of the IELTS as an international benchmarking device, especially as compared to other existing ELT teacher selection criteria instruments; existing theories and literature on this area are limited and fragmented. Qualitative content analysis can be referred to as a research method for "subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh & Shannon, 2005, p. 1278). It is "a method that allows systematic description of the meaning of qualitative material (Schreier, 2012). Qualitative content analysis is a flexible and unobtrusive data collection method used for analyzing all kinds of communication messages (Elo & Kyngas, 2007). It is used in sociology to "develop an understanding of the meaning of communication" and "identify critical processes" (Elo & Kyngas, 2007, pp. 108-9). It is concerned with meanings, intentions, consequences, and contexts (Elo & Kyngas, 2007). Qualitative content analysis is employed "to answer questions such as what, why, and how, and the common patterns in the data are searched for" by using a consistent set of codes to organize text with similar content (Heikkilä & Ekman, 2003, p. 138).

This study can thus benefit from adopting the flexible qualitative content analysis approach in the analysis of the semi-structured interviews and interpreting the different meanings embodied in the informants' different statements. Our aim is to attend to the content meaning of communication in this study and qualitative content analysis has been found suitable for data that require some degree of

interpretation (Schreir, 2012). Qualitative content analysis in this study can help answer what, why, and how questions about Omani ELT educational policy and planning, which the vast majority of studies conducted at SQU so far, have failed to pursue (Al-Issa, 2015).

2.2. Participants

The major source of data in this study came from six STs at the College of Education at SQU. They were selected purposively to help inform the research questions under investigation (Zhang & Wildemuth, 2009). The diverse IELTS scores informed our selection of the six STs. We looked across the scores of the STs who took the IELTS and decided that those six informants would be suitable for the purpose of the study. We had two higher achievers (above Band 6), two lower achievers (below Band 6), and two who matched the Ministry of Education criterion (Band 6).

At the beginning of the semester, an email was sent to all students attending their last semester at SQU, inviting them to participate in the study. In order to encourage participation, the students were informed that their IELTS test fees would be reimbursed. Students who opted to take part filled in an online form confirming their participation and indicating the preferred date/time to be interviewed.

A total of 58 (21 males and 37 females) graduating ELT major students took the academic version of the IELTS. We approached six STs for interviews. It is noteworthy that this study is a part of a large scale research project internally funded by SQU. It investigates Omani SQU ELT STs' language proficiency. The six STs were then divided into three groups: Participants who scored below the overall IELTS Band 6, participants who scored Band 6 overall, and those who scored above Band 6 overall (see Table 1). There were two interviewees in each group.

Table 1. IELTS scores of the interviewed informants

ST. No.	Overall Band Score	Listening	Reading	Speaking	Writing
1	7.5	7.5	8	8	7
2	7.5	7	7	8	7
3	6	5.5	6	6	5.5
4	6	6	5.5	6	5.5
5	5.5	5	5.5	6	5.5
6	5.5	5.5	4.5	5.5	5.5

Guided by Malterud, Siersma, and Guassora's (2015) five-item "information power" pragmatic model, the relatively limited number of informants chosen to be interviewed was based on (a) the narrow aim the study offers, (b) the specificity of experiences, knowledge, or properties among the study informants, (c) the level of established support of theoretical background of the study as seen in the aforementioned reviewed literature, (d) a strong and clear communication between the researchers and informants due to prior knowledge and trust, and (e) the analysis strategy of this exploratory study to present selected patterns relevant for the aims of study.

2.3. Ethical Considerations

We sent the informants a request letter for participation in the study. We clearly stated in the letter the research topic. We also stated that their participation was voluntary and that they had the right to withdraw from the interview at any time they wished. We further made it clear that we would tape the interview session using the voice recording facility in our mobile telephone devices and a digital MP3 recorder and transcribe and analyze them ourselves. In addition, we gave them a protocol letter in which we assured them of the confidentiality of the information they provided.

The six informants agreed to be interviewed after signing a letter of consent vetted by the Deanship of Research at SQU. We fixed an appointment to interview each informant on a separate day. None of the informants complained about the use of the tape recorder or felt that it was intrusive in any way. Each interview was conducted on a separate day after fixing an appointment with the informants. All informants welcomed being interviewed and showed ample flexibility and cooperation.

2.4. Data Collection

Different questions were used in the semi-structured interviews conducted with the informants to stimulate their perceptions about the reliability and validity of the IELTS. Reliability and validity are two aspects upon which empirical tests are usually evaluated. We had the pertinent literature and our longstanding knowledge and experience as Applied Linguists to guide us to ask questions, which would elicit specific answers to those aspects. The following questions were asked.

1. Why did you take the exam in the first place? How important to you is taking the IELTS exam?
2. How important is it to you to know your language proficiency level?
3. Do you think that setting an IELTS score as a pre-condition for working as an English teacher is a fair requirement? Why/ why not? Can you suggest an alternative means of assessing graduates' language proficiency?
4. In your opinion, is the score you obtained a true representation of your language proficiency? Why/why not?
5. What aspects of the test may have affected your score?
6. In your opinion, how may this score in the IELTS exam affect you in the future either positively or negatively?

To improve the quality, reliability, and validity of the interview questions, they were sent along with the study aims to an expert jury, who gave their feedback about some of the items.

All interviews were conducted in English. All informants answered all questions. The interviews varied in length, which we attribute to the different experiences and perceptions of the different informants. We then transcribed the interviews.

Another equally important and substantial source of data was the pertinent literature on the various ELT selection criteria. We used it to help enhance the construct validity of the study (Yin, 2003) and make theoretical contributions to the ELT teachers' selection in the Omani system. Different relevant studies from various parts of the world were found in the literature on the reliability and validity of the IELTS and some of the other alternative selection criteria. Reference to these studies will enhance the argument and the findings in this study and test and add to those earlier theories in the Omani situation (Elo & Kyngas, 2007). The pertinent literature will be used as a comparative guide to the data, which is primarily collected through interviews and should lead to "the identification of presences and absences, as well as similarities and differences" and "is the foundation on which meaning and understanding is based" (Bergman & Coxon, 2005, p. 10).

2.5. Data Analysis

To answer the first, second and third research questions, we used an inductive approach to qualitative content analysis, as we moved from the specific to the general (Elo & Kyngas, 2007). We extracted all the texts from the semi-structured interviews in which our informants talked about the IELTS and the alternative selection criteria and read them several times to familiarize ourselves with the data and make sense of their meanings. As we read the data, and following a manifest content analysis approach, we took down headings in the margins using the text highlighting and review

facility in Microsoft Office Word. The headings we wrote were IELTS, Interviews, and GPA. The last one is complex and represents a Bachelor degree and combines teaching competencies and academic performance, which can be determined through tangible classroom-based outcomes, which are discussed by Freeman et al. (2015) and emphasized by NCATE and ACTFL. One way of determining tangible classroom-based outcomes is through teacher observation. An alternative way is through taking a written exam. Both ways are suggested by some informants and at the end of the day determine STs' GPA.

Because qualitative content analysis is a method for subjective interpretation of the content and systematically describing the meaning of verbal communication messages to develop understanding of the meaning of such messages (Cho & Lee, 2014), we looked for meaning as embedded in the informants' statements and certain key words/terms, which represented characteristics of the reliability and validity sub-categories since none of the informants used direct or explicit technical terms to refer to those sub-categories. The informants' messages communicated different perceptions depending on each informant's experiences, attitudes, and beliefs, which was challenging to us and required rigor and transparency on our part. Guided by Greckhamer and Cilesiz's (2014) approach to discourse analysis, we tabulated "the discourse analysis process to represent the process of analysis and interpretation by providing anchors connecting data units, specific points of reasoning (i.e. concepts), and building blocks" (p. 431) (see Table 2, Table 3, and Table 4). In the first column we included segments of the raw data. The next column included textual units from the pertinent literature. We used the data from the first and second columns to describe the concepts and highlight their main features in the third column. The last column included the name of the code the previous three columns logically led to.

Table 2. Tabulating the discourse analysis process of the IELTS validity

Data Unit in Context	Data Unit	Concept	Building Block
Looking for a job and maybe to complete my higher studies. So some of the universities ask for the IELTS in order to accept you and they need a certain band in IELTS. Some of the local institutions look for an IELTS exam such as the Air Force and the Army. They want the candidate teacher to have a minimum of Band 6 in the IELTS exam (Informant #1).	"Face validity is not really a scientifically based criterion but an impressionistic one" (Faust, 2012, p. 65).	Working as an English teacher in the <i>Air Force</i> has better financial incentives, social prestige and professional opportunities than what is found at school.	Has face validity.

Table 3. Tabulating the discourse analysis process of the IELTS reliability

Data Unit in Context	Data Unit	Concept	Building Block
No! I don't think it is fair because frankly the IELTS test is a speed test and you are challenged very much by the time. Maybe there are learners who are slow but they can produce something accurate even when they are working slowly. Fast is not always correlated with accuracy.	The IELTS does not fit the purpose it is written for. It is unfair, unreliable, and invalid to some candidates with different needs (Freimuth, 2016).	Given the socio-cultural and educational background of the STs, the IELTS is culturally biased and not fair because it is not suitable for all candidates and that due to a lack of familiarity with the test task or its content.	Lacking parallel-forms reliability.

Table 4. Tabulating the discourse analysis process of the alternative reliable and valid selection criteria

Data Unit in Context	Data Unit	Concept	Building Block
Actually, the best way in my opinion is to visit us in schools. Observation I think is the best way. So, you can know the teacher, you can see the teacher in the classroom. This is the most important point here. Because we are teachers, you can see us in the place where we work in. It's all about practice, delivering!	Teachers can develop a sense of authority and expertise as a result of their knowledge of the "local" tasks and responsibilities of their teaching situations and the social and interactional contexts of their classrooms, as represented in "managing the classroom," "understanding and communicating lesson content," and "assessing students and giving them feedback" (Freeman et al., p. 134) while using a global language like English.	English teacher's knowledge is complex and extends beyond merely demonstrating proficiency in the target language. Teaching is complex too and associated with different interrelated theoretical and practical aspects and types of knowledge such as content knowledge, pedagogic knowledge, pedagogic knowledge and support knowledge.	Has criterion-related validity, construct validity, content validity, parallel-forms reliability, and inter-rater reliability.

We also followed a manifest content analysis of the located literature taking the tangible or concrete surface meaning into account. In other words, as we reviewed the literature, we filled each column with the corresponding data we came across from the literature about the IELTS's reliability (see Table 5). This was followed by filling Table 6 about the IELTS' validity and finally filling Table 7 and Table 8 about interviews and classroom observation respectively. This helped us compare and contrast the data we obtained from the interviews and literature and to look for agreements and disagreements and similarities and differences between the two sets of data. The systematic classification process of coding and using a consistent set of codes to organize texts with similar content offered by the qualitative content analysis approach helped us account for all the data using the structured categorization matrix.

Interestingly, as we were coding the data, we found one sub-category, which we felt did not fit the matrix and felt was a different concept. However, guided by certain key words and phrases and based upon the principles of qualitative inductive content analysis (Elo & Kyngas, 2007), we decided to review further literature and could successfully relate those three statements to “systemic” validity, and hence, include them in Table 2.

Table 5.Data coding to structured categorization matrix about how informants’ perceptions on the IELTS validity

	Informants’ Statements	Literature
IELTS Validity	1. Has face validity	1. Has face validity
	2. Has criterion-related validity	2. Lacks criterion-related validity
	3. Lacks content validity	3. Lacks content validity
	4. Lacks systemic validity	4. Lacks systemic validity
	5. Lacks construct validity	5. Lacks construct validity

Table 6.Data coding to structured categorization matrix about how informants’ perceptions on the IELTS reliability

	Informants’ Statements	Literature
IELTS Reliability	1. Lacks parallel-forms reliability	1. Lacks parallel-forms reliability
	2. Lacks inter-rater reliability	2. Lacks inter-rater reliability

Table 7.Data coding to structured categorization matrix about the informants’ perceptions on interviews

	Informants’ Statements	Literature
Employment Interviews	1. Have construct validity	1. Have construct validity
	2. Have content validity	2. Have content validity
	3. Have criterion-related validity	3. Have criterion-related validity
	4. Have parallel-forms reliability	4. Have parallel-forms reliability
	5. Have inter-rater reliability	5. Have inter-rater reliability

Table 8. Data coding to structured categorization matrix about the informants' perceptions on classroom observation

	Informants' Statements	Literature
GPA	1. Has construct validity	1. Has construct validity
	2. Has content validity	2. Has content validity
	3. Has criterion-related validity	3. Has criterion-related validity
	4. Has parallel-forms reliability	4. Has parallel-forms reliability
	5. Has inter-rater reliability	5. Has inter-rater reliability

2.6. Rigor

In this study we adopted a data triangulation approach using two sets of data sources to reduce any academic and professional bias in favor of the English language proficiency selection criteria theories and hence increase credibility and trustworthiness of the findings (Hsieh & Shannon, 2005) and promote confirmability (Shenton, 2004). A substantial amount of the literature reviewed so far has been a discussion of the experiences of various writers in the field.

We converged and diverged the different perceptions as embedded in the informants' statements and the pertinent literature (Hussein, 2009; Riazi, 2016) to help us cancel out our biases about the IELTS as an invalid language proficiency measuring device.

We had to confront our own biases about the selection of the IELTS. We cannot deny that we too had our biases about the IELTS as an inappropriate selection criterion prior to conducting this study, which were more based on personal experiences that lacked academic and scientific rigor. We felt that researching the topic to locate similarities and differences and presences and absences between the literature and our data would help us delineate a clear picture about the reliability and validity of the IELTS and lead us to contribute to the growth of our knowledge as Applied Linguists and to the ELT field, which suffers from limited and fragmented theories about the topic.

3. Findings

3.1. The IELTS Validity

Informants #1 addressed the high face validity of the IELTS as a tool that measures the language proficiency of STs through highlighting some of the professional/economic and academic benefits it can provide respectively.

Looking for a job and maybe to complete my higher studies. So some of the universities ask for the IELTS in order to accept you and they need a certain band in IELTS. Some of the local institutions look for an IELTS exam such as the Air Force and the Army. They want the candidate teacher to have a minimum of Band 6 in the IELTS exam.

Similarly, Informant #6 addressed the face validity of the IELTS. She was more concerned with knowing her proficiency level to recognize her chances of finding a suitable job in an era witnessing continuous challenges and demands imposed by the job market.

Somehow, I know my strength and weakness in language proficiency but maybe, when I apply for a job, that's what will define our language proficiency because they will only take some people and some people will be left around. Somehow we can say, for example, being able to

read and write and communicate in English. I don't know my proficiency level exactly but I know my strengths and weakness.

Unlike the last two informants, Informant #3 was supportive of the IELTS's criterion-related validity. She linked her score in the IELTS with her academic success and improving her listening and writing in which she had scored 5.5, as she considered this an impediment to her professional role as a language source to her students in the future.

It means that I have to improve my language. So how can I be a teacher in the future and teach my students and my level will be 6 on the IELTS. So, I have to improve actually my writing and listening, because I got a lower mark in them. Also, how the students will benefit from me in the future. So, I want to give them the right format of the English language.

Informant #4 additionally addressed the IELTS's criterion-related validity and believed that it could give him confidence as an English language teacher and make him feel a role model to his students, since language knowledge is a fundamental professional requirement for N-NESTs.

I wanted this IELTS to see my score, my level in English. I took it mainly for career reasons. It's very important because I want to see my level. For instance, if I am an English teacher, I have to be professional in my career in order as a teacher when I go to the field, I face many students. It's embarrassing somehow to find a student who is better than you. So it is preferable to develop yourself, to improve your skills, and in the IELTS you will meet these levels.

A different dimension of the IELTS criterion-related validity as a proficiency level predictor was addressed by the next two informants - #2 and #5, both of whom looked at the IELTS as the ultimate internationally recognized English language proficiency indicator and one which defines a candidate's language proficiency and tells him/her about whether s/he needs to develop his/her communicative competence further in an era witnessing speedy spread of English.

I want to have something tangible to tell me how well I am in the language. When I get to know my current level, maybe I can work on myself more. And if my level satisfies me, maybe this will make me learn more. (Informant #2)

Similarly, Informant #5 thought that the IELTS would indicate to him his proficiency level in English. He further showed great belief in the IELTS criterion-related validity.

I think very important. I want to see my efficiency to which degree I have reached and another question I have in my mind that do I want more to improve my language or is it o.k. now, or what I can do for the future to improve. Even I graduated with a good competence of language, I feel I need to improve, because English language doesn't end at a certain point.

Conversely, Informant #3 addressed the content validity of the IELTS. She specifically addressed the writing component.

The writing part, we have different parts, which is describing a diagram, which is a graph. So, we don't know how to describe a graph. It's my first time to see a graph and how to describe it. Actually, I saw in the IELTS book about describing them. I noticed we don't have the skill of describing the graph. So, I had difficulty in that part.

She continued addressing the problematic content validity of the IELTS. This time she turned her attention to the speaking component.

About the speaking part actually we have a topic and we have to talk about it. They asked me about a company in your village, and our village we don't have a company. So, how can I tell her about a company in my village? So, I told her about a company in the city, in Muscat. I found it very difficult because I can't find a company and what's going in these companies. So, I found it very difficult. I didn't know how to talk about it. Talk about companies is not

about my experience, because I encounter them and read them for interest, not to focus on them or how to apply them in my teaching or in the future as well.

Like the last informant, Informant #4 too was critical of the content validity of the IELTS's reading component and attributed his score to sheer luck and blind guessing, rather than linguistic proficiency.

When they are using these skills, special reading, they are not managing. I can promise that with the articles, the reading passages, I didn't read the passages at all. Just doing by guessing and I got 5.5. If you are asking me about the headlines of the passages, I'll tell you I don't know and I actually don't know anything about them. I think it is in the middle, fair or not fair.

In contrast to all the informants discussed above, the next three informants criticized the IELTS for lacking systemic validity. They believed that certain non-linguistic factors affected their overall performance in the IELTS. Informant #4 talked about feeling sick and hungry.

I was sick on the exam day and I was hungry when I got into the speaking, the interview. I was shaking! The first word in my mind I want to say to the interviewee "maybe I'll fall down" because I was very hungry and I was shaking. I didn't eat anything. So I'm just struggling with the speaking.

Similarly, Informant #5, who scored Band 5.5 and 5 on listening and 5.5 on reading thought of the IELTS as lacking systemic validity and generating a strong negative washback effect. However, he considered a different dimension of non-linguistic factors, which was staying out late the night before the exam and feeling the time pressure.

We were at a workshop at the night before the exam and until 12 o'clock. Moreover, I didn't prepare for the exam. Moreover, the time! Being under pressure! I also found it a bit difficult because personally sometimes I require more time to work, to understand. I need more time. When I read or listen to something, my character is analytic person. I require more time to analyze. I spend more time than my classmates.

Informant #6 was critical of the IELTS and saw it as lacking systemic validity too. However, she considered stress and anxiety to be major factors that affected her score.

I was under stress because it was the first time to go and wait for an hour before our name is called in order to go to each room. That created some stress and anxiety. When we went to the room we were calm somehow but the instructions were somehow tough. You don't do this and somehow even they say don't use your pencil. Use this pencil. It created extra pressure.

3.2. *The IELTS Reliability*

The IELTS received its fair share of criticism from the different STs, due to a lack of various aspects of reliability. Informant #1 compared the IELTS to his SQU ELT program and concluded that the IELTS lacked parallel-forms reliability.

Sometimes it's tricky in the IELTS exam. For example, I scored the lowest grade in writing, while I was expecting to get a higher grade. Because I know myself and from my writing courses that I'm good at writing, because my speaking is not that good especially in terms of accuracy and fluency, but surprisingly in the IELTS exam I got the highest grade in speaking and the lowest grade in writing.

Similarly, Informant #3 compared the writing component in the IELTS to the one tested in her SQU ELT program to make her case about the IELTS's lack of parallel-forms reliability.

Actually to be honest with you I'm not happy with 6 and, my writing skill and listening skill as well, because my marks at the Department of English I got at least 8 out of 10 in the writing skill, because all my exams are writing, just writing. So, how it will come 5.5 in the IELTS

test? Actually I decided to do my IELTS again, because I want to see will it be just 6 or above. I think it was a fault because I didn't do well.

Informant #2 additionally criticized the IELTS for being unfair and thus lacking parallel-forms reliability since there was only one form of the academic IELTS, which failed to take different individual abilities into consideration.

No! I don't think it is fair because frankly the IELTS test is a speed test and you are challenged very much by the time. Maybe there are learners who are slow but they can produce something accurate even when they are working slowly. Fast is not always correlated with accuracy.

The same informant then went on to use herself as an example of the inaccurate IELTS score. She quoted her score band and expressed her dissatisfaction with it due to its failure to reflect her genuine level.

I am not fully satisfied with my 7.5 because those who know me said you could have gotten more than that and I had the same view. I am satisfied with it. It is satisfactory, but I could have gotten more.

Moreover, Informant #6, who scored Band 5.5, compared the IELTS to her SQU ELT program overall score, and hence, criticized the IELTS for being unfair due to lacking parallel-forms reliability. She did not see any correlation between the two constructs.

I don't think it is fair because it is a speed test. It doesn't capture the students' ability even though they are active. Because the test has a tight time limit, they will not respond according to their real level. For some students, for example, their GPA is somehow high, but their ability in English is not that much. Maybe they are good in one skill but not in others.

Similarly, Informant #3, who scored Band 6, was critical of the IELTS's lack of parallel-forms reliability. She used her SQU ELT program GPA as an example to corroborate her argument about the IELTS as being an unfair English language proficiency measurement device.

My GPA is higher than my friends' and my friends get more in the IELTS test. So, how they can compare between our grades? For example my grade in the GPA is higher than my friend, but she got on the IELTS test 6.5 and I have 6. So, how can they compare between us? Who will be the teacher in the future?

The same informant then went on to criticize the IELTS speaking component for being marked subjectively, and hence lacking inter-rater reliability through comparing her performance with her friends' to highlight its failure to accurately measure her proficiency at that time.

I will tell you something about the speaking part. Me and my friend, we had the same topic, and with different teachers (examiners). So, most of the students who got that particular teacher got 7 in speaking. Actually, the majority of us who got a different teacher, they got 6.5. The examiners are from the Language Centre. We know them. I communicated with my friends and asked them about what they talked about in speaking, what your topic was about, and what you said about it. So, we discussed our topics and I discovered finally that some of the teachers gave them 7, while the others they got 6, although we have the same level, we share the same ideas, we have the same level of speaking as well, but the most strange thing that teacher gave them 7. How? Why?

3.3. Alternative Reliable and Valid Selection Criteria

Informant #4 accepted the criterion-related validity of the IELTS earlier. However, this time he suggested opting for another more ELT-related selection criterion to help validate the IELTS due to its lack of construct validity.

We have been discussing this with my mates. I asked them, “We are graduating from SQU and we have a bachelor certificate.” So, they don’t care if you got the certificate. But we are focusing on the IELTS. I agree with them that we have to focus on the IELTS but also that means that you are not taking care of our degree; degree of bachelor. It’s important, really important to take the IELTS but not as much as taking a bachelor degree. I think what is important is our bachelor degree. As I told my friends, it is not just a matter of you are good at English.

Similarly, Informant #6 suggested adopting another ELT-related selection criterion to help validate the IELTS. She used her faculty members’ statements to confirm the IELTS’s lack of construct validity.

The employer may give them another exam which might not impact the language skills. But I think if one doesn’t have the IELTS certificate and you don’t employ them, it is not acceptable. And even one professor from the university here told us that why do they give the IELTS test anyway in order to employ you? You are graduates from SQU. You should be employed as soon as possible.

Informant #1 was more precise than the last two informants. He was more in favor of validating the IELTS through incorporating an additional form of benchmarking system with a criterion-related validity, construct validity, content validity, parallel-forms reliability, and inter-rater reliability to help introduce balance to the system. This came in the form of suggesting combining the IELTS score with the results of a series of interviews to judge STs’ proficiency.

It’s very difficult to have a criterion for accepting candidates, but the IELTS could be one option and interviews could be another option. If we combine the two it will be more fair, because mainly this is what most of the institutions now are doing. They ask for the IELTS as a starting point for your application to get accepted, but for you to get accepted you have to go through many interviews.

Informant #3 supported conducting interviews for the reasons outlined by the previous informant. However, she was more explicit about the knowledge tested by the interviewers, which she associated with the content offered by the SQU ELT program.

They can also make interviews with students asking them about what they have taken so far in the English Department. For example, in the English, the Curriculum Systems, about our methods of teaching, in the Department we talk about literature. They can ask about our knowledge, what we have benefited from SQU, the points we are taking so far from the Department. How we can apply all these things in our teaching. How we can apply it, for example, in the future. How you can group students together. How you can teach them specific skills.

Unlike the informants discussed so far, Informant #3 rejected the IELTS as an English language proficiency measurement tool. Instead, she proposed observation, and believed it had criterion-related validity, construct validity, content validity, parallel-forms reliability, and inter-rater reliability. She considered observation as an alternative means of determining STs’ language proficiency and their ultimate solution to full-time enrolment into the Omani ELT force.

Actually, the best way in my opinion is to visit us in schools. Observation I think is the best way. So, you can know the teacher, you can see the teacher in the classroom. This is the most important point here. Because we are teachers, you can see us in the place where we work in. It’s all about practice, delivering!

4. Discussion and Conclusion

This qualitative content analysis study triangulated data from interviews and the pertinent literature about the IELTS and alternative English language proficiency indicator devices to determine the selection of SQU graduates to join the ELT force. The findings indicated that the informants held different perceptions about the reliability and validity of the IELTS, which reflects its failure to serve the needs of the Ministry of Education. Therefore, the Ministry of Education is required to rethink its selection policy, whereby more tangible and classroom-based outcomes are addressed to help and allow for better use of STs' time and energy. In addition, the findings revealed that there were several paradigms and sub-paradigms for reliability and validity, which portrayed a further complex picture about the selection criteria imposed by the Omani Ministry of Education.

The findings showed that the IELTS had high face validity, as stated by Informant #1 and Informant #6, which was in line with the argument forwarded by Hinton (2009), despite the fact that "face validity is not really a scientifically based criterion but an impressionistic one" (Faust, 2012, p. 65). Working as an English teacher in the *Air Force*, as stated by Informant #1, for example, has better financial incentives, social prestige, and professional opportunities than what is found at school.

Additional findings showed that the IELTS had criterion-related validity as found in the statements made by informant #2, #3, #4 and #5, which partly echoed the findings reported by Dooley and Oliver (2002). Such validity positively impacted STs' motivation and confidence and had positive washback through driving instruction and impacting ELT education at the "micro" level. Hence, STs could improve their target language proficiency and the "macro" operation level (Bachman & Palmer, 1996) whereby they could positively impact the BES implementation. Research showed that teachers lacking proficiency in the target language can disturb a challenging CLT-oriented policy implementation like the one implemented in the sultanate (Al-Issa, 2005).

Nonetheless, more findings pointed in the direction of the IELTS as having negative washback effect. This was found in the statements made by informants #3 and #4, who argued against the IELTS's lack of content validity. Informant #4 addressed the reading component, in which he scored 5.5, and criticized its cultural bias and the non-authentic large number of discrete-point item types, which allowed for excessive blind guessing. Such lack of validity thus diminished the learning and acquisition of English as a complex system that required substantial analysis and understanding. It further sent an incorrect message about teacher's communicative competence, which considered all language skills as equally important. Linguistically inadequate teachers like Informant #4, who despite the fact of scoring Band 6, can negatively affect ELT policy implementation due to weakness in one of the four language skills.

Informant #3, on the other hand, was critical of the speaking and writing components. Such findings were in line with the argument forwarded by Hinton (2009) about the IELTS's poor content validity and its cultural bias (Freimuth, 2016). Nonetheless, and in fairness to the IELTS, the examples quoted by Informant #3 were representations of authentic language use, which required demonstrating high-order thinking skills. Nonetheless, the poorly designed and implemented Omani ELT syllabus failed to equip the learners with such fundamental skills (Al-Mahrooqi, 2012a).

Moreover, the characteristics and elements found in the different tests designed at SQU and schools always differed from those found in the IELTS (Al-Issa et al., 2016), which justified the IELTS's lack of "consequential validity" (Messick, 1996) and its failure to drive instruction, as required by the Ministry of Education, to help positively impact ELT policy implementation. It is fair and safe to say, therefore, that there is an obvious lack of alignment between the content and format of the curriculum at SQU and those of the IELTS, which is a "power" test and which is exhausting and requires more time, concentration, different levels of knowledge, and uses advanced reading strategies and higher

order thinking skills (Ying, 2011). Power tests can produce a positive influence on teaching (Messick, 1996). Messick (1994), nevertheless, acknowledged that “it is problematic to claim evidence of test washback if a logical or evidential link cannot be forged between the teaching and learning outcomes and the test properties thought to influence them” (p. 247).

More findings pointed in the direction of the IELTS as having a powerful negative washback effect on STs, due to a lack of systemic validity. This appeared evident in the statements made by informants #4, #5 and #6, bearing in mind that the last two informants failed to achieve Band 6 on the IELTS. Several authors wrote that tests that caused pressure and anxiety to students had a negative washback on learning (Birjandi & Alemi, 2010; Zheng, 2007).

One can argue that a lack of preparation to take the IELTS, which differs greatly from taking other types of exams at Omani school and SQU (Al-Issa et al., 2016), led to generating this kind of negative feelings about the IELTS as being culturally biased. All three informants ascribed their poor score to “locus of control” (Shannon, 2008) – external factors beyond their control, which was an indication of their lack of self-direction and internal control and individual ability to handle changes within their learning environment. One can attribute this lack of internal locus of control to some of the practices found in the Omani ELT system, which directed the students’ effort and ability towards memorizing large chunks of the textbook language, rather than using it creatively and innovatively to achieve multiple significant purposes (Al-Issa, 2014a). The same practices were more or less found in certain parts of SQU too (Al-Issa, 2014b). There is hence a lack of “systemic validity” (Frederiksen & Collins, 1989), or alignment between classroom activities or instruction and assessment, due to the adoption of certain powerful social values and cultural traditions, which lead to a negative washback effect on individual students preparing for a high stakes test, sitting it, and receiving the feedback and decision taken on the basis of the test score (Alderson & Wall, 1993; Bailey, 1996).

The findings in this section reveal that the IELTS suffered from various reliability problems. One of these problems was the lack of inter-rater reliability, as stated by Informant #3, who quoted the speaking component. This partly echoed the findings reported by Veerappan and Sulaiman (2012) and Karim and Haq (2014) and the argument forwarded by Griffiths (2003). The IELTS is a high stakes test, which defines STs’ opportunities in finding a teaching job after graduation. Grading candidates incorrectly or unfairly can thus affect their overall band grade and hence directly affect their chances of being appointed as full-time teachers.

Furthermore, Informants #1, #2 and #3 correlated the IELTS with the language component offered at SQU. Both informants, nevertheless, agreed that the IELTS was unfair in assessing their language proficiency. This finding was in harmony with the findings reported by Al-Malki (2014) and Khemakhem (2016) but contrary to those reported by Bayliss and Ingram (2006). One can argue that focus in the IELTS is more on demonstrating high-order thinking skills and strategies in the four skills and vocabulary using a wide range of materials and task and text types and forms to help accurately measure candidates’ understanding of language proficiency in relation to real-world target domains. However, Al-Mahrooqi (2012b) conducted an exploratory study about the state of communication skills in Oman. It included 58 students representing the different colleges at SQU and regions in the sultanate. She found that English communication skills are either taught indirectly or not taught at all. The author attributed this to poorly trained teachers, which justified the deficits demonstrated by students’ in the Omani school and higher education systems and put the IELTS within a culturally insensitivity perspective.

A different dimension of the IELTS’s lack of parallel-forms reliability was found in the criticism provided by Informants #3 and #6 about the lack of consistency between the IELTS and these STs’ GPAs. This finding was in contrast with the findings reported by Al-Malki (2014), but in line with

those reported by Dooley and Oliver (2002) and Khemakhem (2016). Parallel-forms reliability was thus used to assess the consistency of the results of two tests constructed in the same way from the same content domain. The IELTS and SQU ELT program are obviously a world apart with respect to content, which makes comparison within the context of this study invalid. Put differently, while the IELTS measured language proficiency, SQU ELT program assessed STs in content knowledge, pedagogic knowledge, pedagogic content knowledge, and support knowledge. This type of inconsistency highlighted the complexity of teacher's (language) knowledge on the one hand and the failure of the IELTS to provide a genuine assessment of the STs' language proficiency and their fitness for the job on the other hand. This once again suggested that the IELTS lacked cultural sensitivity.

Informants #4 and #6 thus questioned the aim behind attending SQU ELT program for several years and working hard to achieve a high GPA. This questioning was consistent with the recommendations made by Al-Malki (2014) and the remark made by Khemakhem (2016) about the Ministry of Education reconsidering its overall Band 6 requirements on the IELTS for appointing graduates as full-time English teachers and allowing those who scored below Band 6 on the IELTS to teach, especially that they had spent at least four years learning to become teachers.

Informants #1 and #3 thus proposed interviews and Informant #3 proposed classroom observation as a potential solution that would facilitate STs' full-time appointment. Informant #3 was very specific about the knowledge she would like the Ministry of Education to pay attention to, which in this case is more sophisticated and complex than mere content knowledge and looking at such knowledge in isolation. She advocated paying attention to pedagogic knowledge, pedagogic content knowledge, and support knowledge.

SQU ELT STs are thus trained to serve in Omani schools. By the time they graduate their knowledge about the local context, with particular reference to the CLT-based BES, and due to the diverse theoretical and practical courses they take on- and off-campus during the minimum of four years they spend at SQU, contributes to shaping and enriching their English-for-Teaching knowledge acquisition and accumulation. It further puts them in a position to take teaching responsibility of ELT in Omani schools. Such teacher education program can in turn facilitate achieving internal ELT quality in schools and pave the way for potentially achieving accreditation and international recognition as pursued by SQU.

Therefore, the Ministry of Education can accept the IELTS due to its high face validity and criterion-related validity, with the latter powerfully impacting STs' motivation and confidence and leading to effective and successful BES implementation. However, it has become evident that English teacher's knowledge is complex and extends beyond merely demonstrating proficiency in the target language. Teaching, as discussed by Freeman et al. (2015), is complex and associated with different interrelated theoretical and practical aspects and types of knowledge such as content knowledge, pedagogic knowledge, pedagogic knowledge, and support knowledge, which are mainly the NCATE and ACTFL assessment standards.

The ambitious CLT-based BES national project thus requires teachers to be creative and innovative agents of change and this can be best determined through considering selection criteria like the GPA and structured interviews, due to their reliability and validity as competence indicators that can gauge all aspects of a teacher's knowledge. Embracing such a policy can help reveal how teachers use their theoretical knowledge to inform their practice. It can additionally reveal how those teachers can align their local performances with global set benchmarks to help the Ministry of Education achieve its aims and feed significant knowledge back into SQU ELT program about its structure, content, and implementation in its pursuit of international accreditation and recognition.

5. Limitations and Recommendations for Further Research

Despite the fact that this study has reached its aims, there were some inevitable limitations. First, qualitative content analysis is admittedly time and energy consuming. Next, the context we opted for is very narrow. Finally, this study was conducted on a small size of population. Therefore, it is important to pursue a quantitatively oriented investigation about the effect of the IELTS as a validated selection criterion by the Ministry of Education in the Sultanate of Oman and the role of the alternative selection criteria discussed in this study, or any additional criteria that might be found relevant to the context, to help obtain deeper and broader insights and understanding about this important issue.

Conversely, more qualitative studies can be pursued employing deductive and inductive approaches to content analysis. This is in order to allow for a deeper understanding of the effect and implications of perceptions similar to those discussed in this study or others and as communicated by different informants occupying different positions on the selection criteria of English language teachers in the gulf region and beyond where English and ELT education are centrally planned.

Acknowledgements

We would like to express our deepest appreciation to SQU for funding this internal grant project [IG/LAW/PULW/12/01]. We would also like to express our special thanks to the study participants for providing invaluable data for this study.

References

- Al-Busaidi, K. (1995). English in the labour market in multilingual Oman with special reference to Omani employees. Unpublished doctoral thesis, University of Exeter, England.
- Alderson, J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Al-Issa, A. (2005). An ideological discussion of the impact of the NNESTs' English language knowledge on Omani ESL policy implementation. *Asia EFL Journal On-line*, 7(3). Retrieved from http://www.asian-efl-journal.com/September_05_asmai.php.
- Al-Issa, A. (2014a). Constructing grammar instruction in the Omani EFL system: A critical literacy perspective. *The Qualitative Report*, 19(104), 1-25. Retrieved from <http://www.nova.edu/ssss/QR/QR19/al-issa104.pdf>.
- Al-Issa, A. (2014b). Researching the use of the English language in the Law job market in the Sultanate of Oman: Implications for policy-practice improvement. *Educational Research for Policy and Practice*, 13(1), 25-44.
- Al-Issa, A. (2015). Making a case for new directions in English language teaching research at an Omani university: A critical qualitative content analysis report. *The Qualitative Report*, 20(5), 560-595. Retrieved from <http://nsuworks.nova.edu/cgi/viewcontent.cgi?article=2131&context=tqr>
- Al-Issa, A. & Al-Bulushi, A. (2012). English language teaching reform in Sultanate of Oman: The case of theory and practice disparity. *Educational Research for Policy & Practice*, 11(2), 141-176.
- Al-Issa, A., Al-Bulushi, A. & Al-Zadjali, R. (2016). Arab English language candidates climbing the IELTS mountain: A qualitatively driven hermeneutic phenomenology study. *The Qualitative Report*, 21(5), 848-863. Retrieved from <http://nsuworks.nova.edu/cgi/viewcontent.cgi?article=2367&context=tqr>

- Al-Mahrooqi, R. (2012a). A student perspective on low English proficiency in Oman. *International Education Studies*, 5(6), 263-271. Retrieved from <http://www.ccsenet.org/journal/index.php/ies/article/view/19832>. Doi:10.5539/ies.v5n6p263
- Al-Mahrooqi, R. (2012b). English communication skills: How are they taught at schools and universities in Oman? *English Language Teaching*, 5(4), 124-130. Retrieved from <http://www.ccsenet.org/journal/index.php/elt/article/view/15943>
- Al-Malki, M. (2014). Testing the predictive validity of the IELTS Test on Omani English candidates' professional competencies. *International Journal of Applied Linguistics and English Literature*, 3(5), 166-172. Retrieved from <http://www.journals.aiac.org.au/index.php/IJALEL/article/view/1140>
- Alshammari, M. (2016). IELTS academic reading module test: Validity and reliability. *British Journal of English Linguistics*, 4(2), 26-37. Retrieved from <http://www.eajournals.org/wp-content/uploads/IELTS-Academic-Reading-Module-Test-Validity-and-Reliability1.pdf>
- Arcuino, C. (2013). The relationship between the test of English as a foreign language (TOEFL), the International English language testing system (IELTS) scores and academic success of international Master's students. Unpublished doctoral thesis, Colorado State University, Fort Collins, Colorado.
- Avdi, E. (2011). IELTS as a predictors of academic achievement in a master's program. *EA Journal*, 26(2), 42-49.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, Oxford University Press.
- Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Bayliss, A. & Ingram, D. (2006) IELTS as a Predictor of Academic Language Performance, Australian International Education Conference, pp. 1-12. Retrieved from [http://aiec.idp.com/uploads/pdf/BaylissIngram%20\(Paper\)%20Wed%201630%20MR5.pdf](http://aiec.idp.com/uploads/pdf/BaylissIngram%20(Paper)%20Wed%201630%20MR5.pdf)
- Bergman, M., & Coxon, A. (2005). The quality in qualitative methods. *Forum: Qualitative Social Research*, 6(2), 1-20. Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/457/974>
- Biber D., Gray, B., & Staples, S. (2014). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 36(1), 1-31.
- Birjandi, P., & Alemi, M. (2010). The impact of test anxiety on test performance among Iranian EFL learners. *Broad Research in Artificial Intelligence and Neuroscience*, 1(4), 44-58.
- Cho, J., & Lee, E. (2014). Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *The Qualitative Report*, 19(64), 1-20. Retrieved from <http://www.nova.edu/ssss/QR/QR19/cho64.pdf>
- Clyne, M. (2008). English as an international language: Challenges and possibilities. *Australian Review of Applied Linguistics*, 31(3), 1-10.
- Cranston, J. (2012). Evaluating prospects: The criteria used to hire new teachers. *Alberta Journal of Educational Research*, 58(3), 350-367. Retrieved from <http://ajer.journalhosting.ucalgary.ca/index.php/ajer/article/view/1053/893>

- Dooley, P. & Oliver, R. (2002). An investigation into the predictive validity of the IELTS test. *Prospect*, 17, 36-54. Retrieved from https://www.ielts.org/~media/researchreports/ielts_rr_volume12_report5.ashx
- Elo, S., & Kyngas, H. (2007). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115.
- Faust, D. (2012). *Coping with psychiatric and psychological testimony* (6th ed.). Oxford, Oxford University Press.
- Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, 3(4), 70-85. Retrieved from <http://ehlt.flinders.edu.au/education/iej/articles/v3n4/feast/paper.pdf>
- Fredericksen, J., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), p. 27-32.
- Freeman, D., Katz, A., Gomez, P., & Burns, A. (2015). English-for-teaching: Rethinking teacher proficiency in the classroom. *ELT Journal*, 69(2), 129-139.
- Freimuth, H. (2013). A persistent source of disquiet: An investigation of the cultural capital on the IELTS exam. *International Journal of Education*, 1(1), 9-26.
- Freimuth, H. (2016). An examination of cultural bias in IELTS Task 1 non-process writing prompts: A UAE perspective. *Learning and Teaching in Higher Education: A Gulf Perspective*, 13(1), 1-16.
- Graddol, D. (2006). *English next*. London: British Council.
- Greckhamer, G., & Cilesiz, S. (2014). Rigor, transparency, evidence, and representation in discourse analysis: Challenges and recommendation. *International Journal of Qualitative Methods*, 13, 422-443.
- Griffiths, C. (2003). *Language learning strategy use and proficiency*. Unpublished doctoral thesis, University of Auckland, Auckland, New Zealand.
- Heikkilä, K., & Ekman, S-L. (2003). Elderly care for ethnic minorities—wishes and expectations among elderly Finns in Sweden. *Ethnicity & Health*, 8(2), 135-146.
- Hinton, T. (2009). Retrieved from <http://www.birmingham.ac.uk/Documents/collegeartslaw/cels/essays/testing/THintonTesting.pdf>
- Hsieh, H., & Shannon, S. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288. Retrieved from http://www.iisgcp.org/pdf/glssn/Supplemental_Reading_on_Coding_2.pdf
- Hussein, A. (2009). The use of triangulation in social sciences research: Can qualitative and quantitative methods be combined? *Journal of Comparative Social Work*, (4)1, 2-12. Retrieved from <http://journal.uia.no/index.php/JCSW/article/view/212>
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language proficiency: How distinct? *Applied Linguistics*, 29(1), 24-49.
- Jacob, B, Rockoff, J., Taylor, E., Lindy, B., & Rosen, R. (2016). *Teacher applicant hiring and teacher performance: Evidence from DC public schools*. Cambridge, MA: National Bureau of Economic Research.

- Karim, S., & Haq, N. (2014). An assessment of IELTS speaking test. *International Journal of Evaluation and Research in Education*, 3(3), 152-157. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1091676.pdf>
- Kerstjens, M., & Nery, C. (2000). Predictive validity in the IELTS Test: A study of the relationship between IELTS scores and students' subsequent academic performance. *English Language Testing System Research Reports*, 3, 85-108.
- Khemakhem, S. (2016). Investigating the predictive validity of IELTS for a teacher education program in UAE, (Unpublished doctoral thesis). University of the West of England, Bristol, England. Retrieved from <http://eprints.uwe.ac.uk/28078>
- Klanrit, P. & Sroinam, R. (2012). EFL teacher's anxiety in using English in teaching in the language classroom. *International Journal of Social Science and Humanity*, 2(6), 493-496.
- Malterud, K., Siersma, V., & Guassora, A. (2015). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753-1760. Doi10.1177/1049732315617444
- Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, 19, 203-218.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 13-23.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256. Retrieved from <http://dx.doi.org/10.1177/026553229601300302>
- Moates, J. (2006). Final report. Sultanate of Oman: Ministry of Education.
- Nakata, Y. (2010). Improving the classroom language proficiency of non-native teachers of English: What and how? *RELC Journal*, 41(1), 76-90.
- Pierce, B. (1992). Demystifying the TOEFL reading test. *TESOL Quarterly*, 26(4), 665-691.
- Riazi, M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*. Abingdon, Oxon, England: Routledge.
- Roshan, S. (2013). A critical review of the revised IELTS speaking test. *International Journal of English Language Education*, 2(1), 120-127. Retrieved from <http://dx.doi.org/10.5296/ijele.v2i1.4840>
- Roux, P. (2014). English as an international language: The debate continues. *Polyglossia*, 26, 45-58. Retrieved from http://rcube.ritsumeai.ac.jp/bitstream/10367/5265/1/Peter_W_Roux_English.pdf
- Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: Sage.
- Shannon, S. (2008). Using metacognitive strategies and learning styles to create self-directed learners. *Institute for Learning Styles Journal* 1, 14-28. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi:10.1.1.576.3658&rep=rep1&typepdf>
- Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-75.
- Tajeddin, Z., & Adeb, A. (2016). Native and nonnative English teachers' perceptions of their professional identity: Convergent and divergent. *Iranian Journal of Language Teaching Research* 4(3), 37-54. Retrieved from <http://www.urmia.ac.ir/sites/www.urmia.ac.ir/files/3.pdf>

- Veerappan, V., & Sulaiman, T. (2012). A review on IELTS writing test, its test results and inter rater reliability. *Theory and Practice in Language Studies*, 2(1), 138-143. Retrieved from <http://www.academypublication.com/issues/past/tpls/vol02/01/20.pdf>
- Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney Papers in TESOL*, 1, 51-70. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=807454015B0662080A202CCBE41B2AC?doi:10.1.1.492.6136&rep=rep1&type=pdf>
- Yen, D., & Kuzma J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. *Worcester Journal of Learning and Teaching*, 3, 1-7. Retrieved from <http://eprints.worc.ac.uk/811/1/YenKuzmaIELTScores.pdf>
- Yin, R. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Ying, Z. (2011). IELTS preparation courses in China: The reading comprehension task. Kristianstad University, Sweden. Retrieved from <http://www.divaportal.org/smash/get/diva2:427863/fulltext01.pdf>
- Zhang, Y., & Wildemuth, B. (2009). Qualitative analysis of content. In B. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 308-319). Westport, CT: Libraries Unlimited.
- Zheng, L. (2007). Washback effect of achievement oral test on college English learning—an empirical study. Unpublished master's thesis, University of Guangxi University, Guangxi, China.

Ummanlı İngilizce öğretmenlerinin seçim kriteri algıları: Politika revizyonu için öneriler

Öz

İngilizce yeterliği, ana dili İngilizce olmayan öğretmenlerin mesleğe başlamasında ana kriter olarak belirlenmiştir. Umman Sultanlığı Eğitim Bakanlığı, Sultan Qaboos Üniversitesi mezunlarının tam zamanlı İngilizce öğretmeni olarak alınması için IELTS sınavından 6 alma şartı koymuştur. Bu nitel içerik analizi çalışması, Sultan Qaboos Üniversitesi mezunlarının İngilizce öğretmeni olarak seçimini belirlemek için mülakatlardan ve farklı dil yeterlilik ölçüm araçları ilgili literatürden elde edilen verileri birleştirir. Bulgular Milli Eğitim Bakanlığının IELTS sonuçlarının geçerli bir kriter olarak kabul edilmesini yeniden gözden geçirmesi gerektiğini göstermiştir. Sonuçlar Umman Sultanlığında İngilizce öğretmeni olmak ile ilgili önemli öneriler sunmuştur.

Anahtar sözcükler: İngiliz dili öğretmenliği; aday öğretmen; IELTS

AUTHORS BIODATA

Dr. Ali Al-Issa has a Ph.D. in Education (TESOL) from University of Queensland, Australia. He is currently an Associate Professor of TESOL at the Department of Curriculum and Instruction, College of Education, Sultan Qaboos University. His research interests include foreign/second language education policy and planning and teacher training and education. Dr. Al-Issa has published and presented widely about ELT in Oman. He is currently the best cited Omani researcher.

Dr. Ali Al-Bulushi has a Ph.D. in Applied Linguistics from Lancaster University, UK. He is currently an Assistant Professor of Applied Linguistics at the Department of Curriculum and Instruction, College of Education, Sultan Qaboos University. His research areas of interest include English language teacher professional development, language teacher education, task-based language teaching, computer-mediated communication, and reflective practice.

Dr. Rima Al-Zadjali has a Ph.D. in Applied Linguistics from Lancaster University, UK. She is the Deputy Director for Academic Programs in the Quality Assurance Office, Sultan Qaboos University. Her research areas of interest include language testing, integrated assessment of reading and writing, and academic program quality assurance.